DYNAMIC AND STABLE EPIGENOMIC PROFILES IN MAMMALIAN CELL FATE CONVERSION

A Dissertation

Presented to the Faculty of the Weill Cornell Graduate School

of Medical Sciences

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Ying Liu

August 2015

DYNAMIC AND STABLE EPIGENOMIC PROFILES IN MAMMALIAN CELL FATE CONVERSION

Ying Liu, Ph.D.

Cornell University 2015

A mammalian cell with defined developmental potency can be converted to a pluripotent stem cell that has the potential to differentiate into any of the three germ layers: endoderm, mesoderm, or ectoderm. This reprogramming phenomenon offers a number of potential clinical applications, especially for diseases with a genetic basis. Several strategies have been applied to achieve cell reprogramming, including somatic cell nuclear transfer (SCNT), cell fusion, inducing pluripotent stem (iPS) with overexpression of several key pluripotency-associated transcription factors. However, these current approaches are limited in cell source, reprogramming efficiency, or application safety, which rescrict their potential applications in research and clinical treatment. Besides, the underlying reprogramming mechanisms remain largely unknown. And it is unclear whether intrinsic genetic and epigenetic characteristics of a lineage-restricted cell can affect reprogramming process.

This thesis is focused on the impact of the epigenome on mammalian cell fate conversion from unipotency to multi-/pluri-potency. In Chapter one, we studied the spontaneous conversion from cultured mouse spermatogonial stem and progenitor cells (SSCs) to multipotent adult spermatogonial-derived stem cells (MASCs). In Chapter two, we performed time-series study on transcription factor-induced mouse embryonic fibroblasts (MEFs) reprogramming. In each model system, we applied high throughput sequencing and bioinformatics analyses to dissect both transcriptome and epigenome defined by transcription-associated histone modifications and histone variant exchange.

We found in SSCs that many genes essential to pluripotent stem cell maintenance and differentiation were enriched with both histone H3 lysine 4 and lysine 27 trimethylation modifications (K4me3+K27me3) at promoter regions. After SSC conversion, promoter histone modifications were restrictively changed at pluripotency- or germline-specific genes but remained repressive for most somatic genes, bestowing MASCs with pluripotency-associated promoter chromatin states. At enhancer regions, the core pluripotency circuitry was activated partially in SSCs and completely in MASCs, concomitant with global erasure of germ cell-specific enhancer activity and initiation of an embryonic-like program. In addition, histone variant H3.3, which is specifically enriched at core pluripotency genes in SSCs, was potentially involved in chromatin remodeling and pluripotency gene reactivation during the reprogramming of differentiated MEFs.

These results suggest that unipotent SSCs encode their innate developmental flexibility by means of the epigenome and that both the promoter chromatin state, and the activity of cell-type-specific enhancers are prominent features of SSC reprogramming. Besides, our finding of the gene-specific epigenetic conversion during mammalian cell reprogramming provides insights into the development of new strategies to achieve pluripotency.

BIOGRAPHICAL SKETCH

Ying was born in Beijing, China in 1980. She attended Damochang Elementary School from 1986 to 1992 and Guangqumen High School from 1992 to 1998.

In the fall of 1998, Ying matriculated at Nankai University in Tianjin, China where she majored in Biochemistry. She was honored with university scholarship in 2000 and 2001, and graduated with B.S. degree in 2002.

In the fall of 2002, Ying matriculated at the graduate school of University of Minnesota-Twin Cities, where she majored in Molecular Biology and Genetics. Under the direction of Dr. Deanna Koepp, she completed her thesis in "Analysis of F-box protein function in cell-cycle control system in human cell lines and *Saccharomyces cerevisiae*". She got her Master degree in 2004.

From December 2004 to June 2007, Ying worked as a stuff research associate in Dr. Erin K. O'Shea's group. She relocated with the lab from San Francisco to Boston in the summer of 2005. During this three years, she studied yeast High Osmolarity Glycerol (HOG) MAPK signaling network with multiple systematic methods, including DNA microarray, fluorescent microscopy and programmed signal quantification, etc. She also helped with many publications in Dr. O'Shea's group.

In 2007, Ying decided to pursue advanced education in biology and she matriculated at Weill Cornell Graduate School of Medical Sciences in New York, New York. She studied the mammalian cell reprogramming associated stem cell epigenomics under the supervision of Dr. Shahin Rafii at Weill Cornell Medical College and Dr. C. David Allis at Rockefeller University.

DEDICATION

This work is dedicated to my parents and friends who have always been supportive to my research.

ACKNOWLEDGMENTS

First and foremost I would like to thank my mentors, Dr. Shahin Rafii at Weill Cornell Medical College and Dr. C. David Allis at Rockefeller University, for their guidance, support, and encouragement during my graduate study. This unusual union of mentorships from two top scientists not only has great impact on my understanding about science, but also shapes my life in general. I am also grateful to the members of both the Rafii and the Allis labs. Working in such big labs gets me exposed to different fields of science. Moreover, I have learned to be an active team player and developed many lasting friendships with my labmates.

I would like to thank Dr. Marco Seandel at Weill Cornell Medical College who initiated the SSC epigenome project and generously shared information and experimental materials since I started my graduate research. I also thank Dr. Scott Keeney at Memorial Sloan Kettering Cancer Center for great scientific imput, inspiration, and criticism. I would like to thank Dr. Olivier Elemento at Weill Cornell Medical College for long-term efficient support on bioinformatics. Particularly, his previsou postdoc Dr. Eugenia Giannopoulou, now a faculty member at City University of New York, kindly taught me many programming skills and helped me with prelimiary data analysis. The close collaboration with Dr. Elemento and Dr. Giannopoulou built me a bridge between biology and bioinformatics, and significantly improved my thesis work.

I would also like to thank Dr. Jenny Xiang, the director of Genomics Resources Core Facility at Weill Cornell Medical College, for providing high-quality sequencing service and bioinformatics support.

Finally, I would like to thank all my friends and family who have provided endless love and support for years. You have always been the source of my courage, strength, and passion for living.

TABLE OF CONTENTS

LIST OF FIGURES

Introduction

Chapter One

Chapter Two

LIST OF TABLES

Chapter One

Chapter Two

LIST OF ABBREVIATIONS

ac:	acetylation
A_{diff}:	Differentiated spermatogonia
A_{undiff}:	Undifferentiated spermatogonia
A_s:	Single spermatogonia
AGSC:	Adult Germline Stem Cell
BTB:	Blood-Testis Barrier
ChIP:	Chromatin Immunoprecipitation
ChIP-Seq:	Chromatin Immunoprecipitation and massively parallel Sequencing
DSB:	Double-Strand Break
ESC:	Embryonic Stem Cell
FACS:	Fluorescence-Activated Cell Sorting
GDNF:	Glial cell line-Derived Neurotrophic Factor
H3K27:	H3 lysine 27 residue
H3K36:	H3 lysine 36 residue
H3K4:	H3 lysine 4 residue
HFSC:	Hair Follicle Stem Cell
q/a-HFTAC:	quiescent/activated-Hair Follicle Transient-Amplifying matrix Cell
HMT:	Histone Methyltransferase
HSC:	Haematopoietic Stem Cell
ICM:	Inner Cell Mass
IF:	Immunofluorescence
iPS:	induced Pluripotent Stem cell
KDM:	Histone lysine Demethylase
LIF:	Leukemia Inhibitory Factor
MASC:	Multipotent Adult Spermatogonial-derived stem Cell
me1:	monomethylation
me2:	dimethylation
me3:	trimethylation
MEFs:	Mouse Embryonic Fibroblasts
MSCI:	Meiotic Sex-Chromosome Inactivation
PGC:	Primordial Germ Cell
PHD:	plant homeodomain
PTM:	Post-Translational Modifications
PRIM:	Promoter ChIP-seq Read Intensity ratio for histone Modification
PS:	Pachytene Spermatocyte
RD:	Replication Dependent
RI:	Replication Independent
RS:	Round Spermatid
SCNT:	Somatic Cell Nuclear Transfer
SSC:	Spermatogonial Stem and progenitor Cell
TNP:	Transition Nuclear Protein
TSS:	Transcription Start Site

INTRODUCTION

I. Spermatogonial Stem and Progenitor Cells (SSCs)

1. SSCs specification, development, and differentiation

a. Embryonic origin of SSCs

Mammalian spermatogonial stem and progenitor cells (SSCs) are the only germline stem cells that differentiate to all subsequent germ cells in the adult male gonad. In mice, primordial germ cells (PGCs), the embryonic precursors of SSCs, originate from the post-implantation epiblast cells between embryonic day (E) 6.0 and 6.5, and establish a cluster of about 45 cells at E7.25 (Figure 1) (Ginsburg et al., 1990). Upon germline specification, somatic genes are mainly repressed in PGCs (Saitou et al., 2002), while expression of several embryonic stem cells (ESCs) signature transcription factors (e.g., Pou5f1/Oct4, Nanog, Sox2, Lin28a, etc.) are reactivated (Pesce et al., 1998; Yamaguchi et al., 2005; Yabuta et al., 2006; Kurimoto et al., 2008; West et al., 2009). PGCs are therefore distinct from somatic cells with repression of somatic transcriptome and expression of ESCs signature genes, which retrieve pluripotent developmental potency in this cell population. Subsequently, PGCs migrate into the male genital ridge between E10.5 and E11.5, proliferate in the embryonic gonad till E13.5, and enter G1-phase mitotic arrest as gonocytes (prospermatogonia) after E13.5 (Figure 1). Many PGC-specific genes are activated in the genital ridge, including Ddx4, Sycp3, Dazl, etc. (Sasaki and Matsui, 2008).

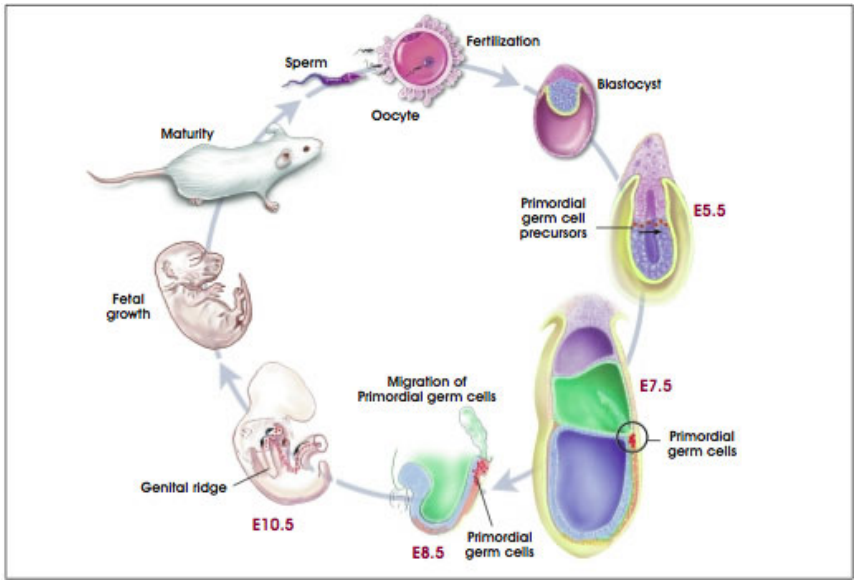


Figure 1. PGC specification and migration in post-implantation embryo. (Adapted from 2001 Terese Winslow, Caitlin Duckwall)

b. Spermatogonial self-renewal

Within the first week after birth, some mouse gonocytes resume proliferation and mature into single spermatogonia (A_s), which are morphologically distinct from their cell division progenies in that they do not have heterochromatin and intercellular bridges (Figure 2A) (de Rooij and Russell, 2000). These A_s spermatogonia locate in the basal lamina of the seminiferous tubules, form the first wave of spermatogenesis and establish the initial pool of SSCs (Figure 2B) (Oatley and Brinster, 2006). In adult mouse testes, A_s spermatogoina was estimated to represent about 0.03% of germ cells by whole mounts of seminiferous tubules (Tagelenbosch and De Rooij, 1993). However, only about 10% of A_s spermatogoina functioned as SSCs in transplantation, presumably dues to the low efficiency of passage through the blood-testis barrier (BTB) to the stem cell niche (Nagano et al., 1999). Indeed, SSC homing was shown to be more efficient in immature pup testis that lacks BTB tight junctions between Sertoli cells (Shinohara et al., 2001).

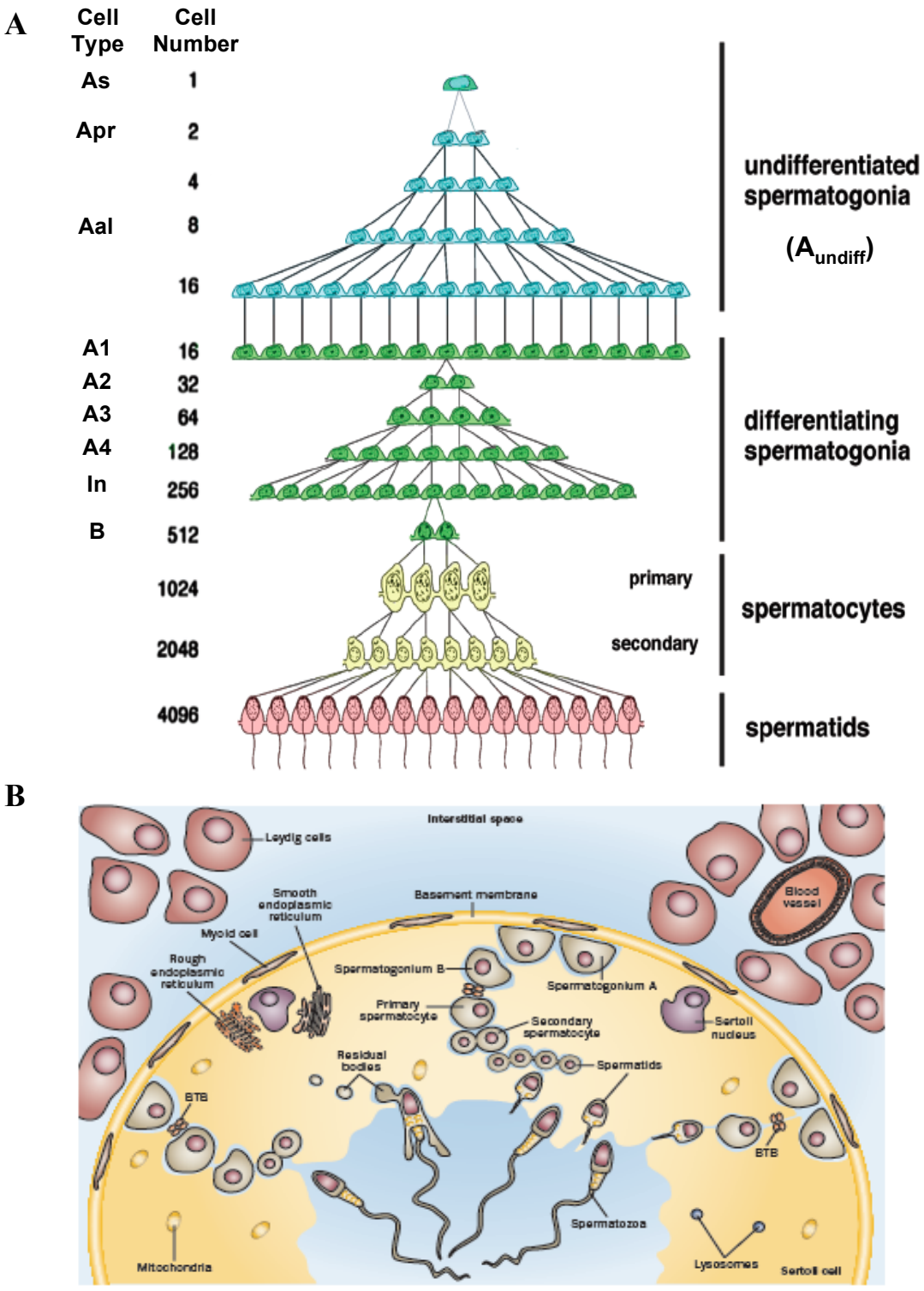


Figure 2. Germ cell development in mice.

(A) Spermatogenesis in postnatal mouse testes. (Adapted from (Yoshida, 2008))
(B) Schematic illustration of mammalian spermatogenesis. (Adapted from (Rato et al., 2012))

SSCs undergo self-renewal in the niche determined by Sertoli cells (Figure 2B) (Oatley et al., 2011b). As the only somatic cells that directly interact with germ cells, Sertoli cells not only provide structural support but also growth factors essential for the maintenance of SSC self-renewal. The key Sertoli cell-secreted growth factors include glial cell line-derived neurotrophic factor (GDNF) (Meng et al., 2000) and fibroblast growth factor 2 (FGF2) (Ishii et al., 2012), both can support SSC self-renewal and proliferation *in vivo* and *in vitro*.

Besides growth factors, several transcription factors are involved in SSC self-renewal (Phillips et al., 2010; Kanatsu-Shinohara and Shinohara, 2013). Studies with knock-out mouse models suggested that Zbtb16/Plzf (Buaas et al., 2004; Costoya et al., 2004), Etv5 (Chen et al., 2005; Morrow et al., 2007), Taf4b (Falender et al., 2005), Id4 (Oatley et al., 2011a), and Bcl6b (Oatley et al., 2006) function in a cell-autonomous manner to maintain SSC self-renewal, and the mutant mice progressively lose spermatogenesis in part or whole testes. Notably, the expression of several ESC signature transcription factors (i.e., Pou5f1/Oct4 (Pesce et al., 1998), Sox2 (Arnold et al., 2011), Lin28a (Zheng et al., 2009), etc.) are also confined to the undifferentiated spermatogonia, although their roles in SSC development are not well characterized (Seandel et al., 2010). Conversely, SSCs do not express genes that promote embryonic somatic cell development, which is similar to their progenitor PGCs in post-implantation embryo.

The A_s spermatogonia (including SSCs) also express several stem cell-specific cell surface antigens. For example, mouse testicular cells with α6- and β1-integrin preserve stem cell activity (Shinohara et al., 1999), and SSCs were found nearly exclusively in the Thy-1^+c-kit^- cells (Kubota et al., 2003). Besides, Gfrα1 (Meng et al., 2000) and

GPR125 (Seandel et al., 2007) were also reported as potential SSCs marks in mammals.

c. Postnatal development and differentiation of SSCs

As unipotent stem cells, SSCs undergo both self-renewal and differentiation throughout adulthood. The fate decision of a SSC to produce differentiating progeny initiates spermatogenesis in postnatal mouse testes (Figure 2A) (Oatley and Brinster, 2006). During this process, A_s spermatogonia (including SSCs) take a series of mitotic cell divisions to generate both undifferentiated spermatogonia (A_{undiff}) and differentiating spermatogonia (A_{diff}), which then undergo meiotic cell divisions to generate spermatocytes and haploid spermatozoon (Figure 2A) (Yoshida, 2008). The A_{diff} spermatogonia are distinguishable from A_{undiff} spermatogonia on the basis of many characteristics including being c-kit positive (Shinohara et al., 1999; Ohta et al., 2003). Correspondingly, study with mutant mouse has clearly demonstrated that c-kit expression is a distinctive and critical step of the spermatogonia differentiation (Yoshinaga et al., 1991). However, no known SSC-specific cell surface antigen has been found to distinguish SSCs from other types of A_{undiff} spermatogonia.

From sexual maturity onward, A_{diff} spermatogonia undergo mitotic division to form primary spermatocytes (spermatocytes I), which then activate meiotic differentiation and divide two times into diploid secondary spermatocytes (spermatocytes II) and then haploid round spermatids. During post-meiotic spermatogenesis, nucleosomal histones in round spermatids are firstly replaced by basic transition nuclear proteins (TNPs) and then by protamines. This histone-to-protamine exchange packages the genome in extremely compact formation and yields mature spermatozoa, which are the transcriptionally inert final product of

spermatogenesis. Theoretically, 4096 spermatozoa can be produced from a mouse A_s spermatogonium after 12 cell divisions. And one spermatogenesis cycle takes 35 days to complete in mice (Russell et al., 1990).

The continuous cell differentiation process generates highly heterogeneous germinal population in adult testes. Because the meiosis process induces extensive variations in DNA content and chromatin structure, germ cells at multiple stages of spermatogenesis can be analyzed and isolated by flow cytometry using Hoechst 33342 (Ho) dye, which diffuses through cell plasma membrane and binds with high affinity to poly(d[AT]) sequences in the DNA minor groove. Ho staining methodology can efficiently discriminate preleptotene spermatocytes (meiosis prophase I), spermatocyte I (diploid), spermatocyte II (haploid), round and elongated spermatids (haploid). Basing on the efflux of Ho, a side population (SP) that was suggested to contain SSCs can also be isolated by flow cytometry (Figure 3) (Bastos et al., 2005).

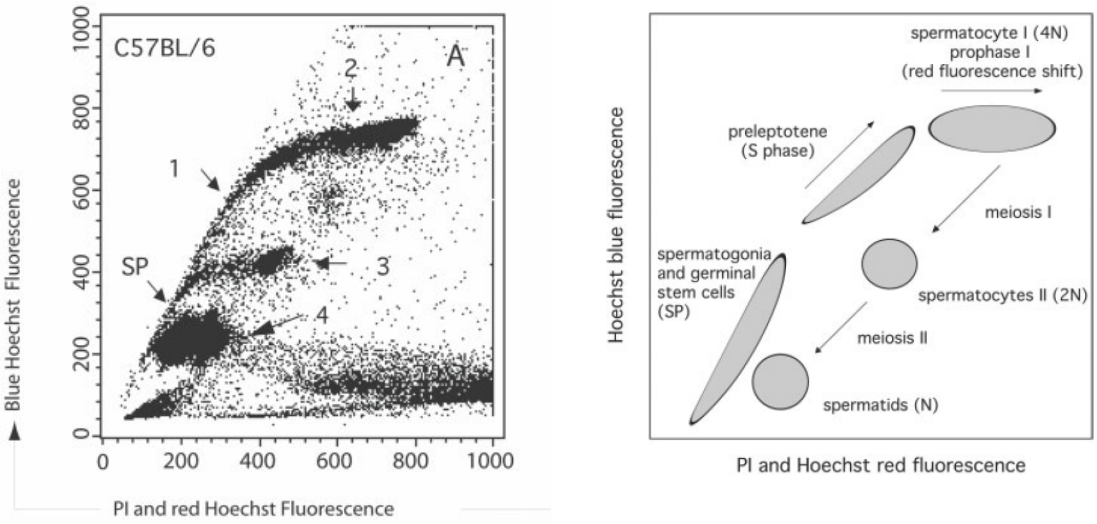


Figure 3. Flow cytometric analysis of Hoechst 33342 stained adult mouse testicular cells. 1, preleptotene spermatocytes (at last S phase before meiosis. DNA content: between 2c and 4c). 2, spermatocyte I (at meiosis prophase I. DNA content: 4c). 3, spermatocyte II (at post-meiosis I. DNA content: 2c). 4, spermatid (at post-meiotic. Hapolid DNA content: 1c). SP, side population with premeiotic spermatogonial including SSCs. (Adapted from Figure 1A (Bastos et al., 2005))

2. Characteristics of SSC culture

a. Isolation and long-term expansion of SSCs

The long-term expansion of mouse SSC culture provides a key tool for studying SSC biology at the cell and molecular level *in vitro*. To establish a culture, an SSC-enriched cell population can be collected from whole testicular cells by several methods, including use of spermatogenesis-defective experimental cryptorchid donors (Shinohara et al., 2000a), isolation by cell surface antigens (Shinohara et al., 1999; Kubota et al., 2003; 2004a; 2004b), and differential plating (Shinohara et al., 2000b). SSCs typically show grapelike morphology on feeder cells and can be passed for years without losing fertility (Figure 4A) (Kanatsu-Shinohara et al., 2003). Besides, A_{undiff} spermatogonia specific cell surface antigens α6-integrin and Thy-1 are stably expressed in 80 – 90% of long-term expanded cells (Figure 4B). Notably, as it has been reported that all types of A_{undiff} spermatogonia share similar cell surface antigens, SSCs have not been able to be specifically isolated from A_{undiff} spermatogonia. As a result, SSC culture is a heterogeneous population with SSCs and other types of A_{undiff} spermatogonia (Kanatsu-Shinohara and Shinohara, 2013). It was reported that the fluorescence-activated cell sorting (FACS) isolated Thy-1^+c-kit^- cells contains nearly all the SSCs with about 6% purity (1 in 15 cells is an SSC) (Kubota et al., 2003). And only about 1–2% of long-term cultured SSCs were estimated to exhibit testicular repopulation capacity (Kanatsu-Shinohara et al., 2005b). However, many groups also reported that even c-kit^+ differentiating germ cells that presumably lose stem cell activity can convert back to SSCs *in vivo* (Barroca et al., 2009) and *in vitro* (Nakagawa et al., 2007). Thus, SSC frequency in culture could be much higher than previously estimated (Kanatsu-Shinohara and Shinohara, 2013).

To support stem cell self-renewal and survival without promoting the growth of

contaminating somatic cells, SSCs need to be expanded in a serum-free or serum-low (1%) condition (Kubota et al., 2004a; Kanatsu-Shinohara et al., 2005a). Moreover, many growth factors have been examed for their ability to promote SSC expansion *in vitro*, including GDNF, FGF2, epidermal growth factor (EGF), leukemia inhibitory factor (LIF), and stem cell factor (SCF, kit ligand) (Kubota et al., 2004a). Among them, GDNF plays a critical role to maintain and stimulate SSC self-renewal in long-term culture (Kubota et al., 2004b). The central importance of GDNF on SSCs expansion *in vitro* well reflects its function in mouse spermatogonia development *in vivo* (Meng et al., 2000), and has been validated in many other mammalian species (Hamra et al., 2005; Sadri-Ardekani et al., 2009). Based on these findings, Kanatsu-Shinohara et al. developed a long-term SSC culture system with GDNF, FGF2, EGF, and LIF. With or without mouse embryonic fibroblasts (MEFs) as feeder cells, SSCs derived from both neonatal and adult mouse testes can stably proliferate in this medium for more than two years and retain spermatogenesis capability (Kanatsu-Shinohara et al., 2003; 2004; 2005a). Although LIF is suggested to promote the survival of gonocytes in newborn testicular cell culture, it is dispensable for SSC self-renewal. And SSC culture can be established from pup or adult testes without LIF (Kanatsu-Shinohara et al., 2007). Study in Dr. Rafii's lab also established a novel $CD34^+$ testicular stromal cell line (JK1) that facilitates long-term SSC culture. And SSCs maintained on either MEFs or JK1 share similar biology characteristics (Kim et al., 2008). Therefore, I derived and expanded adult mouse SSCs in the medium designed by Kanatsu-Shinohara et al. without LIF, and used JK1 or MEFs as feeder cells.

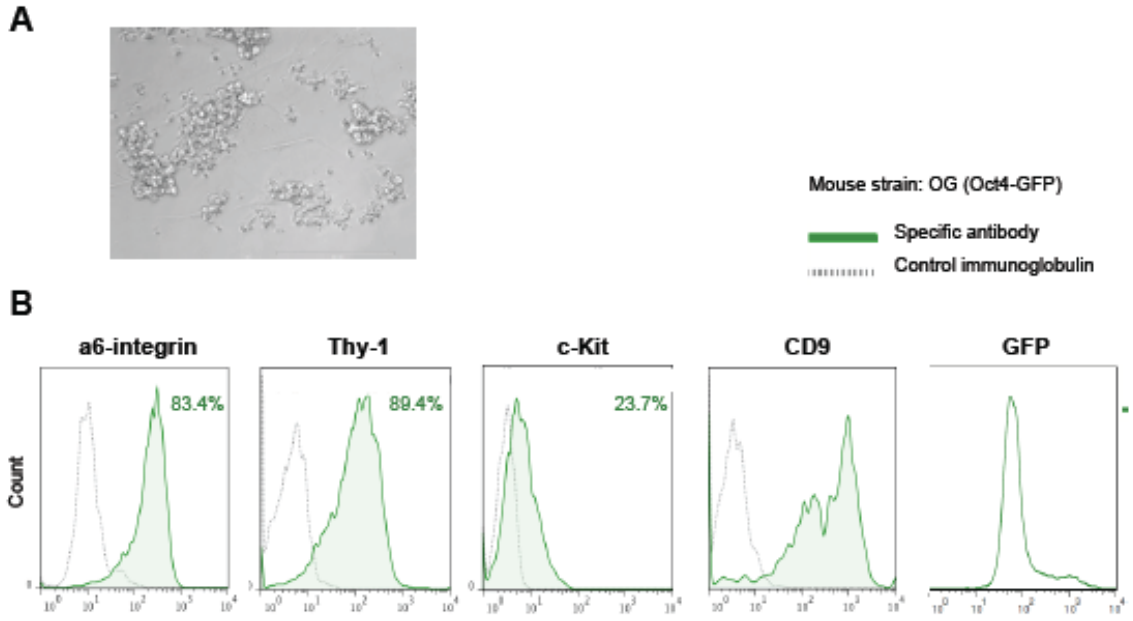


Figure 4. Long-term SSC culture preserves spermatogonia specific characteristics.

(A) Mouse SSC culture on MEFs.
(B) Flow analysis of spermatogonia specific cell surface antigens.

b. Evaluation of stem cell characteristics in SSC culture

To evaluate spermatogenesis and germline transmission capabilities of *in vitro* expanded SSCs, a spermatogonial transplantation technique was developed. The technique involves microinjection of dissociated donor testicular cells into the seminiferous tubules of germ cell-depleted or -deficient recipient testes. After injection, SSCs within the donor cells migrate and colonize stem cell niches on the basement membrane, repopulate, reinitiate spermatogenesis, and produce offspring with the donor haplotype (Figure 5). This donor-derived spermatogenesis produces mature spermatozoon after 35 days, corresponding to one spermatogenesis cycle *in vivo*. This result further confirms that the donor SSCs but not differentiating

spermatogonia contribute to the spermatogenesis (Brinster and Avarbock, 1994; Brinster and Zimmermann, 1994). At present, spermatogonial transplantation is the only functional SSC assay that not only directly validates the presence of SSCs but also investigates their biological characteristics as unipotent germline stem cells *in vivo* (Oatley and Brinster, 2006). Besides, the transplantation assay confirmed that SSCs increased their number during *in vitro* proliferation.

The cultured SSCs expressed both embryonic and spermatogonial specific genes as A_{undiff} spermatogonia *in vivo*. SSCs in long-term culture express many ESC-signature transcription factors, including Pou5f1/Oct4 (Kanatsu-Shinohara et al., 2005b; Seandel et al., 2007; Dann et al., 2008), Klf4 (Kanatsu-Shinohara et al., 2008), Lin28a (my data). However, the expression level of these pluripotency-associated genes is lower than ESCs by either mRNA or protein, and no mRNA or functional proteins were identified for Sox2 (Imamura et al., 2006; Seandel et al., 2007) and Nanog (Kanatsu-Shinohara et al., 2004; Seandel et al., 2007) in cultured SSCs. SSCs also express many well-characterized spermatogonia marks, including Zbtb16/Plzf, Etv5, Id4, Bcl6b, GPR125, Gfrα1, Pou3f1/Oct6 (Wu et al., 2010), Piwil4 (Carmell et al., 2007), etc. (Seandel et al., 2007). However, genes that regulate embryonic differentiation to somatic lineages are mostly silenced in cultured SSCs, for example, Hox family, Prmt8, etc. (see Chapter I).

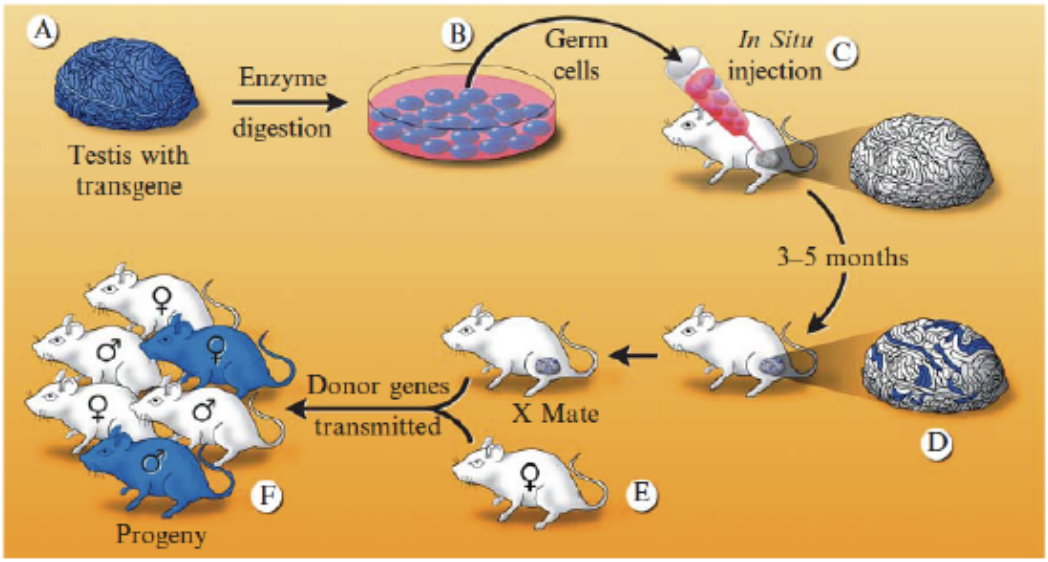


Figure 5. The SSC transplantation technique in mice.

The donor cells with a germ cell-expressed reporter transgene are collected from long-term culture or male testes (A) and digested to become a single-cell suspension (B). These cells are subsequently microinjected into the seminiferous tubules of an infertile germ cell-depleted recipient (C). Donor SSCs can colonize and undergo spermatogenesis in recipient testes. After 3 – 5 months, colonies that express the donor cell-carried reporter transgene can be detected in the recipient testes (D). Each donor-derived colony is generated from a single transplanted SSC. After mating the recipient male to a wildtype female (E), offsprings with the donor haplotype can be produced (F). (Adapted from Figure 1 (Oatley and Brinster, 2006))

3. SSC reprogramming to multipotency

a. Potential pluripotency of SSCs

As the precursors of germ cells, mammalian SSCs undergo unipotent differentiation in adult male gonad, while still possess ultimate totipotent developmental potency to propagate across generations. One example of the unique association between germ cells and pluripotency is the spontaneous formation of teratomas in strain 129 mice (Stevens and Little, 1954). Teratomas contain tissues derived from all three germ layers, suggesting that germ cells in both postnatal and adult testes retain

developmental potency unrestricted to germline. Besides, direct trans-differentiation of spermatogonia into somatic lineages has also been observed upon changing the cell microenvironment (Boulanger et al., 2007) or using a tissue recombination methodology (Simon et al., 2009).

The expression of key pluripotency regulators and repression of somatic genes in both A_{undiff} spermatogonia *in vivo* and SSC culture *in vitro* clearly show that unipotent SSCs and pluripotent ESCs share many similarities in transcriptomes. Pou5f1/Oct4, together with Sox2 and Nanog, form the core pluripotency circuitry in ESCs to maintain stem cell self-renewal and to control the expression of many differentiation genes (Boyer et al., 2005; Loh et al., 2006). It is not known whether the modest level of Pou5f1/Oct4 expression exerts similar or partial function in pluripotency regulation in SSCs.

b. Multipotent adult spermatogonial-derived stem cells (MASCs)

Several groups reported that multipotent adult spermatogonial-derived stem cells (MASCs) can be derived during *in vitro* expansion of mouse unipotent SSCs (Figure 6). MASCs share many common features with pluripotent ESCs, including capabilities to differentiate into various somatic cells *in vitro*, to induce teratomas *in vivo*, and to form germline chimera after blastocyst injection (Figure 6) (Kanatsu-Shinohara et al., 2004; Guan et al., 2006; Seandel et al., 2007; Izadyar et al., 2008; Kossack et al., 2009). MASCs do not contribute to spermatogenesis after being transplantated to seminiferous tubules, suggesting that they lose the capability to differentiate to germ cells (Kanatsu-Shinohara et al., 2008). Notably, SSCs isolated from both neonatal and adult testes are capable of this spontaneous conversion to multipotency (Kanatsu-Shinohara et al., 2004; Seandel et al., 2007). To date, this is the only known

reprogramming event that converts unipotent adult stem cells back to a near pluripotent state without delivery of exogenous genes or gene products, which distinguishes it from induced pluripotent stem (iPS) cell formation (Takahashi and Yamanaka, 2006). However, the spontaneous and sporadic SSC conversion process takes more than 4 weeks to generate multipotent MASCs, and the reprogramming efficiency can be 100-fold lower than generating iPS from MEFs (Kanatsu-Shinohara et al., 2004; Seandel et al., 2007). Many methods were suggested to improve SSC conversion, including deriving SSCs from neonatal or *p53* KO mice (Kanatsu-Shinohara et al., 2004), isolating an unique germ cell population by SSC-specific marks (Seandel et al., 2007; Izadyar et al., 2008), initiating culture with low cell density (Ko et al., 2009), or applying LIF and glycogen synthase kinase-3 (GSK-3) inhibitor CHIR99021 in culture (Moraveji et al., 2012).

MASCs share similar characteristics with ESCs in their differentiation potency both *in vitro* and *in vivo*. Many studies showed that MASCs can be differentiated to functional neurons (Glaser et al., 2008), mature hepaticphenotype (Loya et al.), cardiomyocytes (Baba et al., 2007), and endothelial cells (Seandel et al., 2007), etc. Interestingly, MASCs were reported to express higher levels of mesodermal marks than ESCs (Seandel et al., 2007). *In vitro* differentiation assay also showed that MASCs are more efficient in generating mesodermal cells comparing to ESCs, suggesting that un-equivalent differentiation preferance toward mesodermal lineage could be a specific characteristic in MASCs (Baba et al., 2007).

These observations indicate that intrinsic genetic and epigenetic features rather than enforced expression of ectopic transcription factor(s) are responsible for reprogramming of SSCs. However, the underlying mechanisms of SSC conversion remain largely unknown.

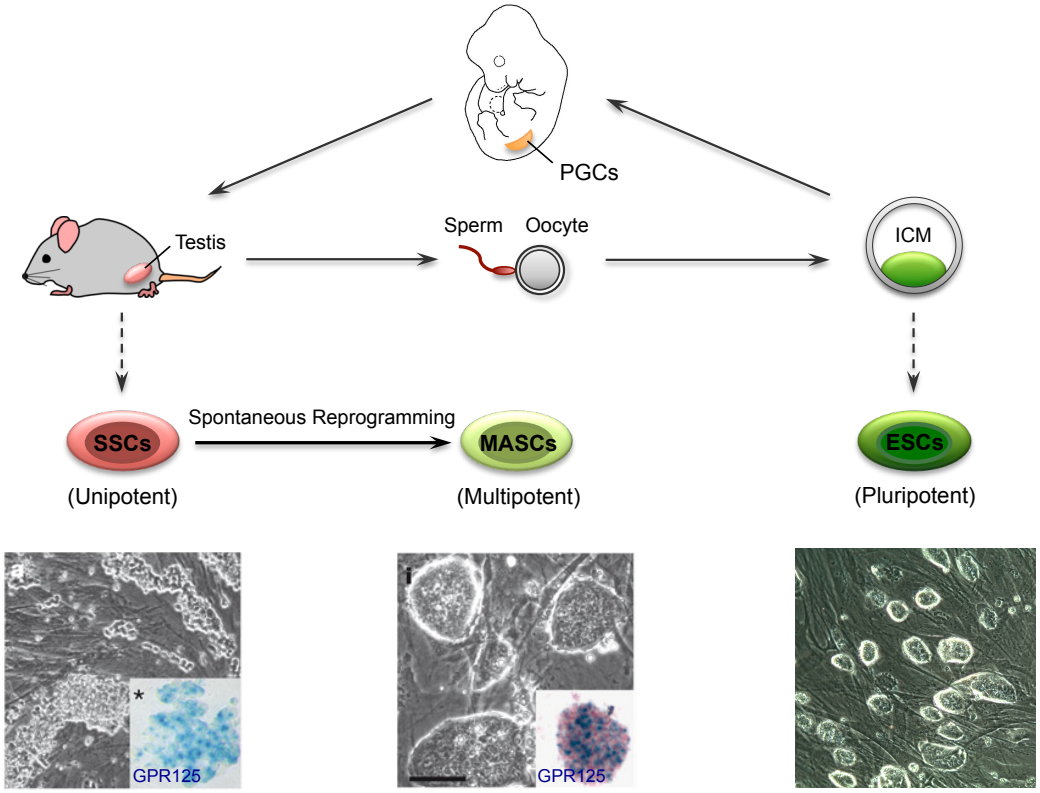


Figure 6. SSCs convert to MASCs.

Unipotent SSCs (red) derived from adult mouse testes and expanded *in vitro* are capable of spontaneous reprogramming to multipotent MASCs (light green), which share similar stem cell characteristics with pluripotent ESCs (dark green) that are derived from inner cell mass (ICM) in the preimplantation embryo. Images of *in vitro* cultured SSCs and MASCs are adapted from (Seandel et al., 2007).

II. Epigenetic Regulation of Transcription

1. Histone modification

In eukaryotic cells, DNA is tightly packed with proteins to form chromatin, which has nucleosome as its basic subunit. Nucleosome is composed of an octamer of core histone proteins (two copies of H2A, H2B, H3 and H4) wrapped with ~146 base pairs of DNA (Figure 7A). As one of the most stable protein-DNA complexes under physiological conditions, nucleosome organization directly influences almost all DNA-related processes including replication, repair and transcription. This unique character makes nucleosome an important regulator in eukaryotic cell development (Kouzarides, 2007).

Nucleosomes are not static, they are instead dynamically regulated by many binding proteins. Among them, histones contribute most of the diversity through their post-translational modifications (PTMs) and variants (Hake and Allis, 2006; Ruthenburg et al., 2007). Because histone PTMs and variants may change the chromatin character and regulate gene expression without changing the underlying DNA sequence, they are proposed to be "epigenetic" factors that regulate eukaryotic cell diversity. And "epigenetics" is defined as "the study of any potentially stable and, ideally, heritable change in gene expression or cellular phenotype that occurs without changes in Watson-Crick base-pairing of DNA" (Goldberg et al., 2007).

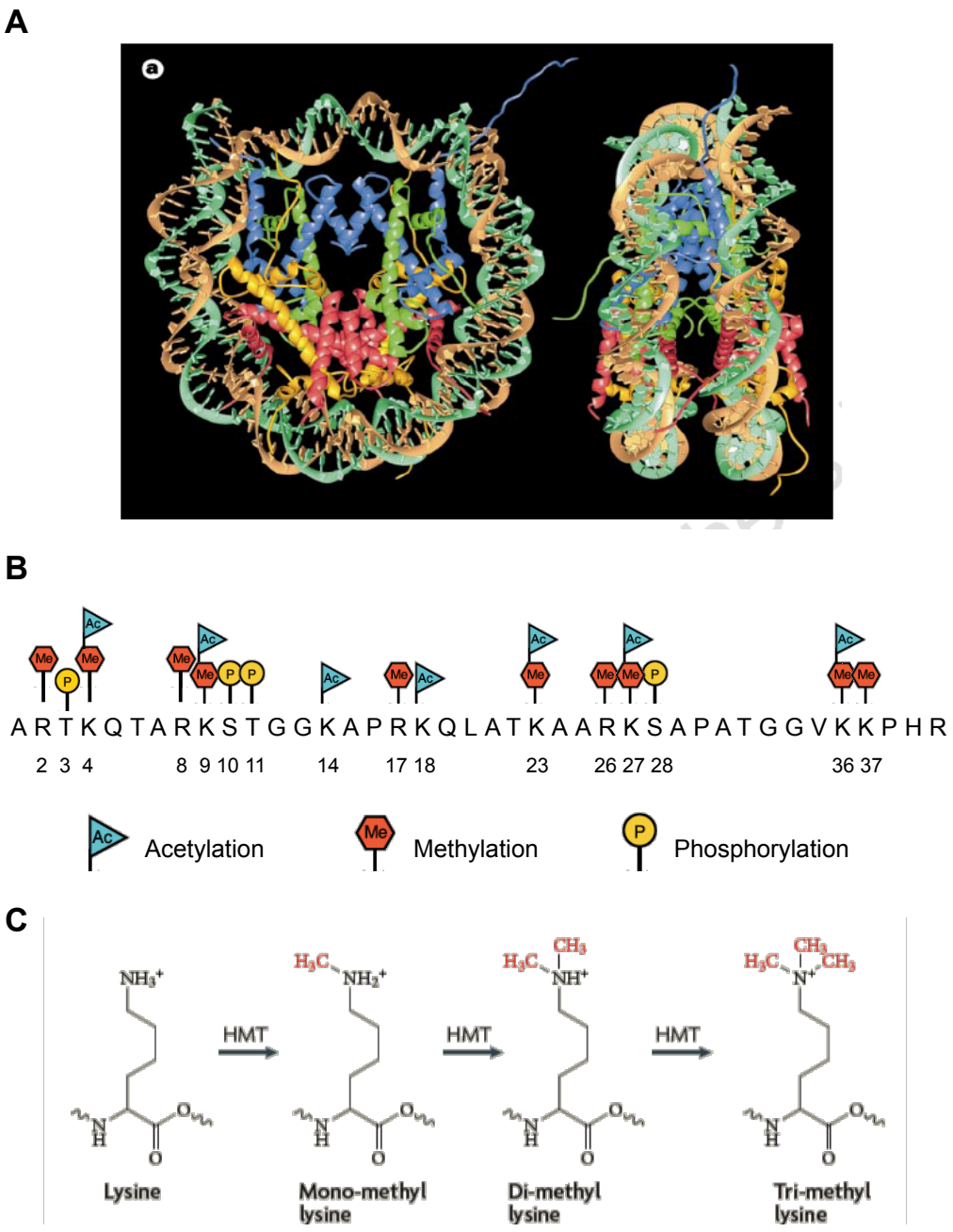


Figure 7. Nucleosome, histone, and histone PTMs.

(A) Nucleosome core particle. Left, a view down the DNA superhelix axis. Right, a view perpendicular to the DNA superhelix axis. Brown and turquoise, 146-bp DNA phosphodiester backbones. Blue, H3. Green, H4. Yellow, H2A. Red, H2B. (Adapted from Figure 1a (Luger et al., 1997))

(B) Histone PTMs in the N-terminal tail of canonical H3. Blue flag, acetylation. Red hexagon, methylation. Yellow circle, phosphorylation. Number under each amino acid, position in the sequence.

(C) Sequential methylation on lysine.

a. Histone post-translational modifications *(PTMs)*

The four core histones (H2A, H2B, H3 and H4) are relatively similar in structure and size, and their sequences are highly conserved from yeast to human. Their C-terminal regions form stable globular structures and are deeply buried within the nucleosome. But the unstructured N-terminals protrude from the nucleosome particles and form flexible "tails" (Figure 7A).

Histones regulate nucleosome plasticity mainly through the PTMs on their amino acid residues. Many types of modifications have been identified, including methylation, acetylation, phosphorylation, ubiquitination, sumoylation, etc. Most of the PTMs locate on the residues within the N-terminal "tails" (Figure 7B)(Ruthenburg et al., 2007). Many PTMs have been found to be dynamic, and enzymes that specifically modify different residues have been well identified. Proteins and domains that preferentially associate with different PTMs have also been intensively studied. Histone PTMs provide the chromatin diversity by regulating the contacts between nucleosomes, as well as by recruiting various protein cofactors to chromatin. Both can affect cell development through many DNA-related processes, including transcription (Kouzarides, 2007).

b. Lysine methylations & enzymes

Among all the known PTMs, lysine methylation is one of the best studied modifications. In eukaryotes, lysine (K) can be sequentially methylated to three forms: monomethylation (Kme), dimethylation (Kme2), or trimethylation (Kme3) (Figure 7C). Different lysine residues can be modified in each histone tail, each form of modification (mono-, di-, or tri-methyl) on a specific lysine residue may couple with a special cellular event and process a unique function (Shilatifard, 2008).

Many site-specific lysine methyltransferases (HMTs) and demethylases (KDMs) have been identified. For example, H3 lysine 4 residue (H3K4) can be specifically trimethylated by MLL/WRD5 complex (Wysocka et al., 2005) and demethylated by JARID1/KDM5 family proteins (Christensen et al., 2007; Iwase et al., 2007; Yamane et al., 2007). H3 lysine 36 residue (H3K36) can be methylated by NSD1 and demethylated by Jhdm1b (Rayasam et al., 2003; He et al., 2008). The polycomb repressive complex 2 (PRC2) is responsible for the (di- and tri-) methylaiton of H3 lysine 27 residue (H3K27) through its enzymatic subunits Ezh1 (Shen et al., 2008) and Ezh2 (Margueron et al., 2008; Margueron and Reinberg, 2011), whereas UTX and JMJD3 can erase the methylation from K27 (Agger et al., 2007; Lan et al., 2007).

c. Histone PTMs & transcriptional regulation

Methylation on the residues H3K4 and H3K36 are implicated in activation of transcription. In yeast, H3K4 trimethylation (K4me3) is established together with transcription elongation and requires association with RNA Polymerase II (RNAP II) (Li et al., 2007). But these modifications do not affect transcription elongation or processivity of RNAP II (Mason and Struhl, 2005). However, H3K36 trimethylation (K36me3), a modification established by Set2 following elongation, can specifically recruit Eaf3 and mediate deacetylation of newly deposited histones. This process is proposed to erase the transcription elongation-coupled histone acetylation and avoid intragenic transcription initiation (Carrozza et al., 2005). Our knowledge in yeast PTMs leads us to propose their similar functions in higher eukaryotes. However, mammalian core histones are normally encoded in multi-copied tandem clusters, which add great challenge to genetic manipulation (Sierra et al., 1982). As a result, the

precise function of many PTMs in mammalian cell differentiation and organ development still remains myserious.

Besides of active marks described above, trimethylation on H3K9 (K9me3) and H3K27 (K27me3) are associated with transcriptional repression. Methylation at H3K9 is implicated in the heterochromatin formation and silencing of enchromatic genes with the recruitment of HP1 to the promoters. K27me3 modification has been implicated in the silencing of HOX gene and imprinting genes (Kouzarides, 2007).

Recently, genome-wide studies in different eukaryotic organisms showed that methylations on many lysine residues are tightly regulated and significantly associated with transcription (Figure 8). For example, K4me3 is specifically enriched around transcription start site (TSS) of active genes, while K36me2/3 and trimethylation on H3 lysine 79 residue (K79me3) only exist within the body of transcribing genes. These localization patterns are well conserved from yeast to human (Bernstein et al., 2005; Pokholok et al., 2005). Furthermore, these modification levels are positively correlated to the transcription frequency (Pokholok et al., 2005). Besides, H3K4 mono-/di-methylation (K4me1/2) and H3K27 acetylation (K27ac) enrichments at cis-regulatory DNA elements associate with the activation of these enhancers, which recruit transcription factors, RNAPII and chromatin regulators to positively influence transcription at distal promoters (Zhou et al., 2011). Conversely, the repressive marks K9me2/3 and K27me3 are often found in large broad domains in differentiated mammalian tissues, but K27me3 is more focal at developmental gene promoters in pluripotent ESCs (Wen et al., 2009; Hawkins et al., 2010). It will be interesting to know whether these PTMs are merely markers passively added during transcription, or they can play positive roles in regulating mammalian gene transcription.

Indeed, accumulating evidences reveal the key regulatory functions of histone PTMs in gene expression. Because K4me3 is a mark for active transcription, it is proposed to provide signal for transcription initiation and transcription frequency (Wysocka et al., 2005; Li et al., 2007). This hypothesis is supported by the identification of many K4me2/3 specific binding proteins, which all contain plant homeodomain (PHD) finger and are proposed to interpret K4me3 mark to multiple transcriptional events (Zhang, 2006). These PHD finger proteins are associated with many transcription cofactor complexes, including chromatin-remodeling factor NURF (BPTF) and histone-modification complex mSIN3/HDAC (ING2) (Barak et al., 2003; Shi et al., 2006).

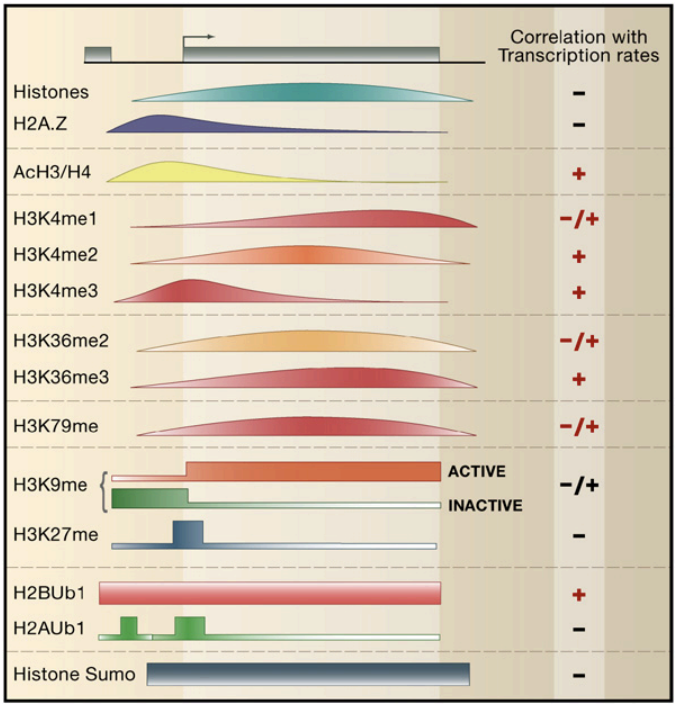


Figure 8. Genome-wide distribution pattern of transcription-associated histone modifications. The distribution of histones and their modifications are mapped on an arbitrary gene relative to its promoter (5' end), gene body, and 3' end. With the exception of the data on K9 and K27 methylation, most of the data are based on yeast genes. (Adapted from Figure 1 (Li et al., 2007))

d. Histone PTMs & cell development

Studies in yeast showed that mutations in H3K4, H3K36, or H3K79 result in relatively little growth defects. But the triple mutant on all the three residues (H3K4R / H3K36R / H3K79R) is lethal by mitotic cell cycle delay and progressive transcription defect (Jin et al., 2007). Although the constitutive association between HMTs and mutated H3 may partially contribute to the lethality, the potential function of these PTMs in transcriptional regulation and cell growth can still not be excluded.

More recently, genome-wide study in mammalian ESCs unveiled a unique association between H3 lysine methylation and ESC-specific gene expression. In ESCs, trimethylation modifications on both H3 K4 and K27 residue are found to co-exist at many promoters and evolutionarily conserved sequences. These areas of chromatin are called "bivalent domains", because K4me3 is a mark for activating chromatin, while K27me3 is a well-known repressive chromatin mark (Bernstein et al., 2006). K4me3+K27me3 bivalent modifications are normally found in developmentally regulated genes, which are silenced in pluripotent ESCs but are activated upon differentiation. As a result, a combination of both K4me3 and K27me3 modifications are proposed to poise genes in a "transcription-ready" state (Figure 9). This hypothesis brings an interesting connection between PTMs coupled transcriptional regulation and mammalian ESC development. But direct evidence is still missing to support this model. The precise function of these lysine methylations in mammalian transcriptional regulation and cell development still remains unknown.

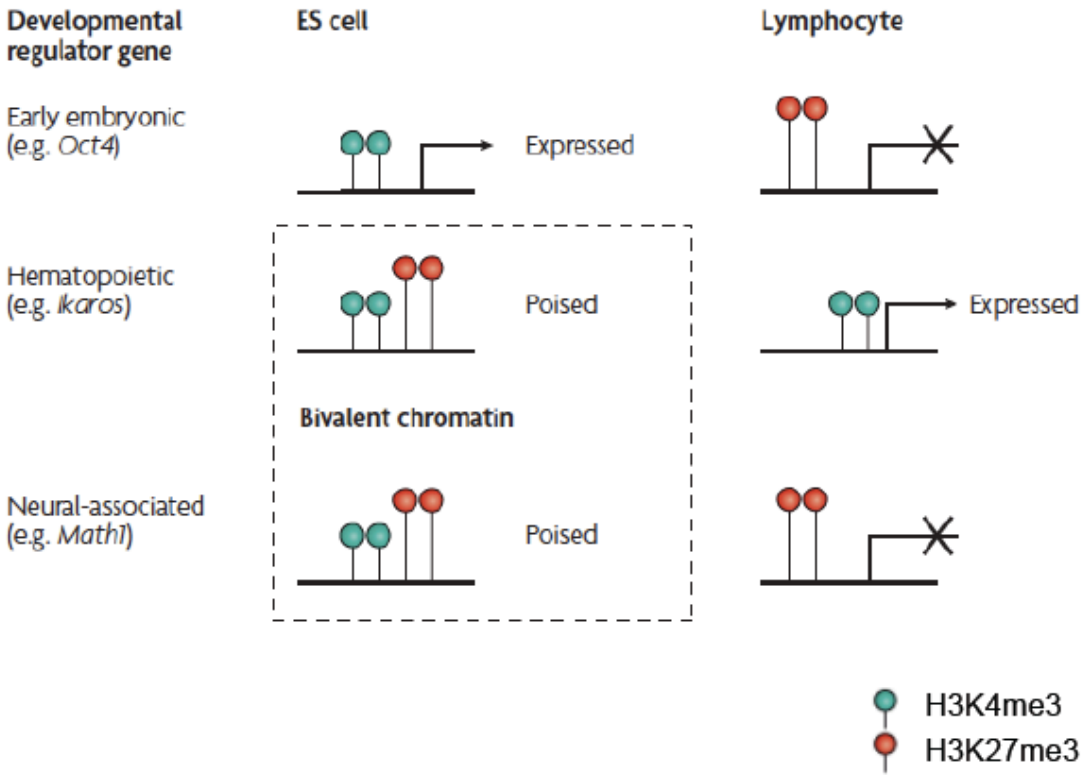


Figure 9. Bivalent chromatin profiles in ESCs. In mammalian ESCs, the promoters of many non-transcribed developmental genes bear a combination of 'conflicting' histone modifications that are normally associated with either active chromatin states (H3K4me3, green) or inactive chromatin states (H3K27me, red). This indicates that these genes are 'poised' for expression in response to appropriate developmental cues. During differentiation, 'bivalent' chromatin profiles are generally resolved, leading to transcriptional activation of tissue-specific genes and silencing of loci associated with alternative developmental pathways. (Adapted from Figure 1 (Spivakov and Fisher, 2007))

2. Histone variant

Except for H4, all the other core histones have variants. They are mostly encoded by single-copy genes located outside the histone gene clusters, and the protein sequences are very similar to their canonical counterparts. Interestingly, many histone variants have unique expression patterns, deposition mechanisms and genome localizations. The variations on a few critical residues are also proposed to cause nucleosome structural heterogeneity. Growing evidence is supporting the notion that histone variants, as histone PTMs, also contribute to the chromatin diversity and regulate many DNA-related events. (Pusarla and Bhargava, 2005)

a. H3 variants (H3.3)

Except for a centromere-specific H3 variant CENP-A, mammals have three somatic H3 variants, H3.1, H3.2, and H3.3 (Figure 10A). Among them, H3.1 and H3.2 are exclusively expressed and assembled into nucleosome during S phase in a replication-dependent (RD) manner, while H3.3 is expressed in both proliferating and quiescent cells and is deposited in a replication-independent (RI) manner (Ahmad and Henikoff, 2002; Szenker et al., 2011). Two independent chaperone systems facilitate H3.3 incorporation into chromatin, with H3.3-specific chaperone HIRA responsible for deposition into protein-coding regions (Tagami et al., 2004; Goldberg et al., 2010), and the ATRX/DAXX complex responsible for deposition at repeat elements like telomeres and pericentric heterochromatin (Figure 10B) (Drané et al., 2010; Goldberg et al., 2010; Lewis et al., 2010).

The “N-terminal tail” region in H3.3 is nearly identical to H3.1 and H3.2, with only one amino acid difference (Serine 31). Although this region is necessary for the deposition through RD pathway, it is dispensable for H3.3 specific RI deposition.

Comparing with H3.2, H3.3 also has three residues difference in the C-terminal structured region, which has been shown to be critical for its deposition through RI pathway (Figure 10A) (Ahmad and Henikoff, 2002). However, it is still unclear whether these differences in H3.3 can bring any structural diversity to chromatin.

Unlike the other histone proteins, mammalian H3.3 is expressed from only two genes: H3f3a (H3.3A) and H3f3b (H3.3B) (Wellman et al., 1987). Gene trapping studies in mouse showed that H3.3A gene is ubiquitously expressed during early embryonic development and in many adult organs, and disrupting H3.3A gene results in partial neonatal lethality while surviving mutants display neuromuscular deficits (Couldrey et al., 1999). Expression of H3.3B gene is more tissue-specific. Targeted disruption of H3.3B gene in mouse is reported to be semilethal, and the surviving males are completely infertile with defects at different stages of spermatogenesis (Bush et al., 2013; Yuen et al., 2014).

To distinguish H3.3 from the other H3 variants, our lab has generated mice that carry a HA epitope tag at the C-terminal of endogenous H3f3b gene (H3.3-HA) (Wen et al., 2014a; 2014c). Heterozygous H3.3-HA mice are viable and fertile. However, homozygous H3.3-HA mice are postnatally lethal with significantly low birth weight (Wen et al., 2014c). The cause of lethality is still under investigation.

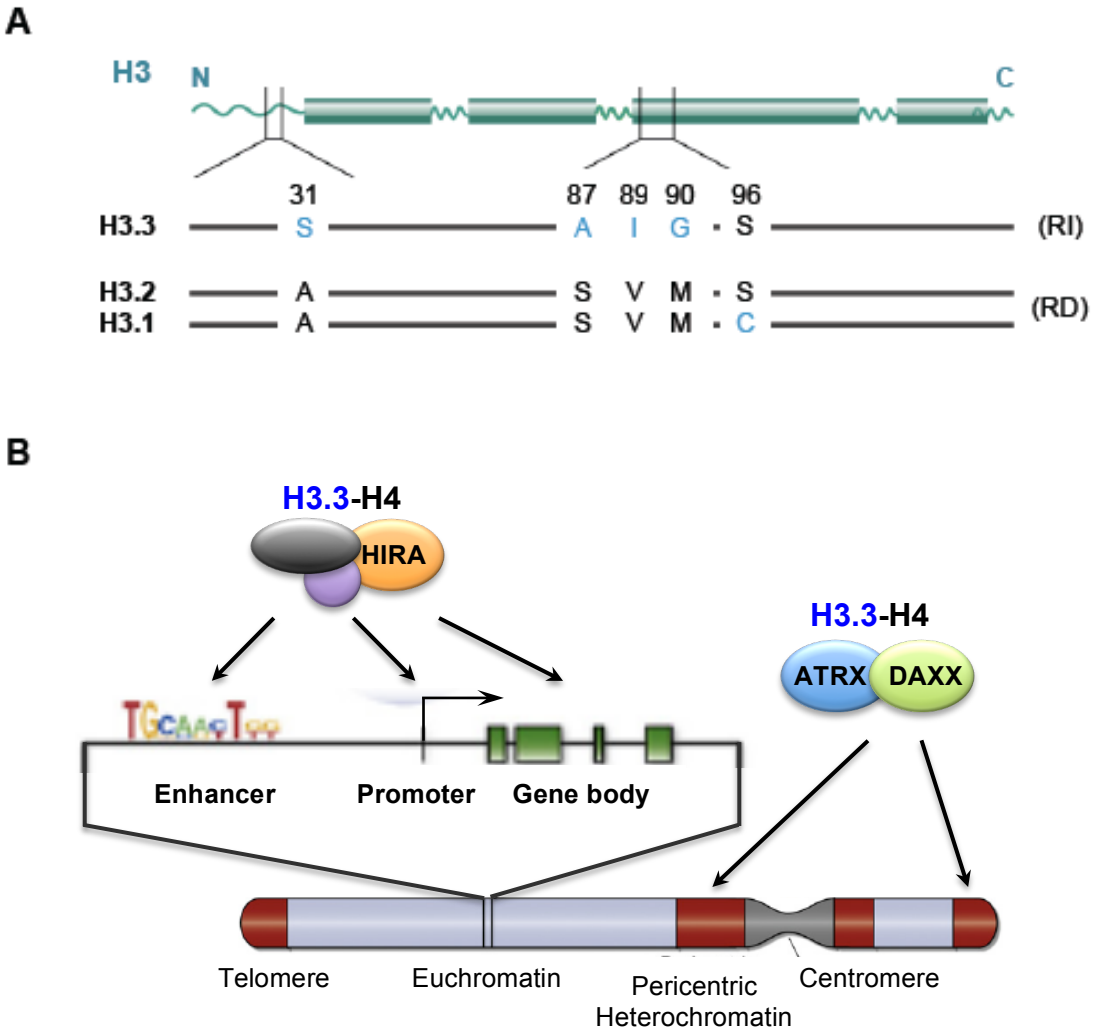


Figure 10. H3 variants in somatic and embryonic mouse cells.

(A) H3 variants. The amino acid differences and corresponding positions at the N-terminal tail and the C-terminal structured regions are highlighted. RI, replication-independent. RD, replication-dependent.

(B) H3.3 enrichment and deposition pathways. In euchromatin, H3.3-specific HIRA complex is responsible for H3.3 deposition in the body of active genes and at promoters regardless of gene activity. In heterochromatin, the ATRX/DAXX complex is responsible for H3.3 deposition at pericentric heterochromatin and telomeres (ESC-specific). The chaperone complex that mediates H3.3 deposition at regulatory elements remains to be identified. (Modified from (Banaszynski et al., 2010))

b. H3.3 & Transcriptional Regulation

Global analyses of *Drosophila* and mouse genomes showed that H3.3 is highly incorporated at the promoters, within the actively transcribed gene bodies, on the boundary of cis-regulatory elements, and at the telomere (Figure 10B) (Mito et al., 2005; 2007; Wong et al., 2009). Some studies also reported that H3.3 is enriched in PTMs associated with active transcription, for example, K4me2, K4me3 and K36me2; but is deficient in PTMs marking gene silencing (McKittrick et al., 2004; Loyola et al., 2006; Garcia et al., 2008). As a result, H3.3 is proposed to be a marker of active transcription in euchromatin regions (Hake and Allis, 2006). However, H3.3 region-specific assembly may also cause structural changes in local chromatin, which may regulate corresponding gene expression and many other DNA-related events. This possibility is waiting to be confirmed or excluded.

c. H3.3 & cell reprogramming

H3.3 has been implicated in chromatin remodeling and epigenetic regulation of gene expression during various cell reprogramming processes (Banaszynski et al., 2010).

In the early mouse zygote, maternal H3.3 plays an important role in male pronucleus formation (van der Heijden et al., 2005). The asymmetrical incorporation of H3.3 into the paternal genome is observed before the transcription activation in both the paternal and maternal pronuclei (Torres-Padilla et al., 2006). Recent studies also suggest that H3.3 deposition into the paternal genome is required for the enrichment of K27me3 modification and the establishment of pericentric heterochromatin, which ensure proper chromosome segregation during the first mitosis (Santenard et al., 2010).

Study in somatic cell nuclear transfer (SCNT) embryos also suggests that H3.3, but not H3.1/H3.2, is a maternal "reprogramming factor" that is essential for chromatin reorganization in the donor nucleus, resulting in pluripotency gene reactivation and reprogramming of a differentiated cell into the pluripotent state (Jullien et al., 2012; Wen et al., 2014a).

Accumulating evidences suggest that H3.3 can assist mammalian stem cell differentiation and embryonic development. In mouse ESCs, H3.3 is suggested to be required for proper establishment of bivalent chromatin modification through faciliating PRC2 recruitment to developmental gene promoters. As a result, several lineage-specific genes are misregulated with reduced levels of K27me3 in H3.3-depleted ESCs, resulting in alteration of the differentiation potency of ESCs (Banaszynski et al., 2013). It consists with the result that, although knocking out H3.3 specific chaperon HIRA in mouse ESCs gives no observable effect, HIRA-/- mutant causes embryonic lethal at the gastrula stage (Roberts et al., 2002).

III. Epigenetic Change in Mammalian Germ Cell Development

1. Genome-wide epigenetic changes during PGC differentiation

It is well known that PGCs undergo extensive epigenomic changes during their migration and early residence within the embryonic gonad. Immunohistochemistry study showed that at the time of PGC specification in post-implantation embryo (E7.25), several transcriptional repression-associated epigenetic marks are similar between PGCs and somatic cells (Seki et al., 2005; 2007). Subsequently, migrating PGCs erase genome-wide DNA methylation and H3K9me2 and then acquire high level of H3K27me3 by E9.5, which presumably complements the erasure of previous epigenetic marks to maintain a proper repressive chromatin state in PGCs (Figure 11) (Seki et al., 2007; Sasaki and Matsui, 2008). The global changes of transcriptional repression-associated epigenetic marks potentially play an important role in suppressing the somatic gene expression programme in migrating PGCs. Besides, the reprogramming on germ cell epigenome might restore and transmit totipotency into the zygote (Sasaki and Matsui, 2008).

Extensive epigenetic reprogramming also occurs in post-migratory PGCs, as reflected by the genome-wide progressive erasure of parental imprints (E10.5 – E12.5) and establishment of paternal imprints (after E14.5), which are then maintained throughout the male germ cell development (Figure 11) (Davis et al., 1999; Ueda et al., 2000; Li et al., 2004; Oakes et al., 2007). The epigenomic changes have been suggested to be important in activating PGC-specific genes (i.e., Ddx4, Sycp3, Dazl, etc.) in the genital ridge (Sasaki and Matsui, 2008).

Several recent studies applied chromatin immunoprecipitation and massively parallel sequencing (ChIP-Seq) to interrogate genome-wide epigenetic marks in PGCs at different developmental stages. Interestingly, K4me3+K27me3 bivalent histone modification was found at the promoters of many developmental regulatory genes in post-migratory PGCs (E11.5 and E12.5) (Lesch et al., 2013; Ng et al., 2013; Sachs et al., 2013). Transcription of these epigenetically 'poised' genes is repressed in PGCs but activated in somatic lineages to direct a wide range of cell differentiation. Furthermore, a significant percent of these PGC bivalent genes retain stable K4me3+K27me3 histone modification in both male and female germ cells following the sex determination (E13.5) and female meiosis (E14.5) (Lesch et al., 2013; Ng et al., 2013; Sachs et al., 2013)

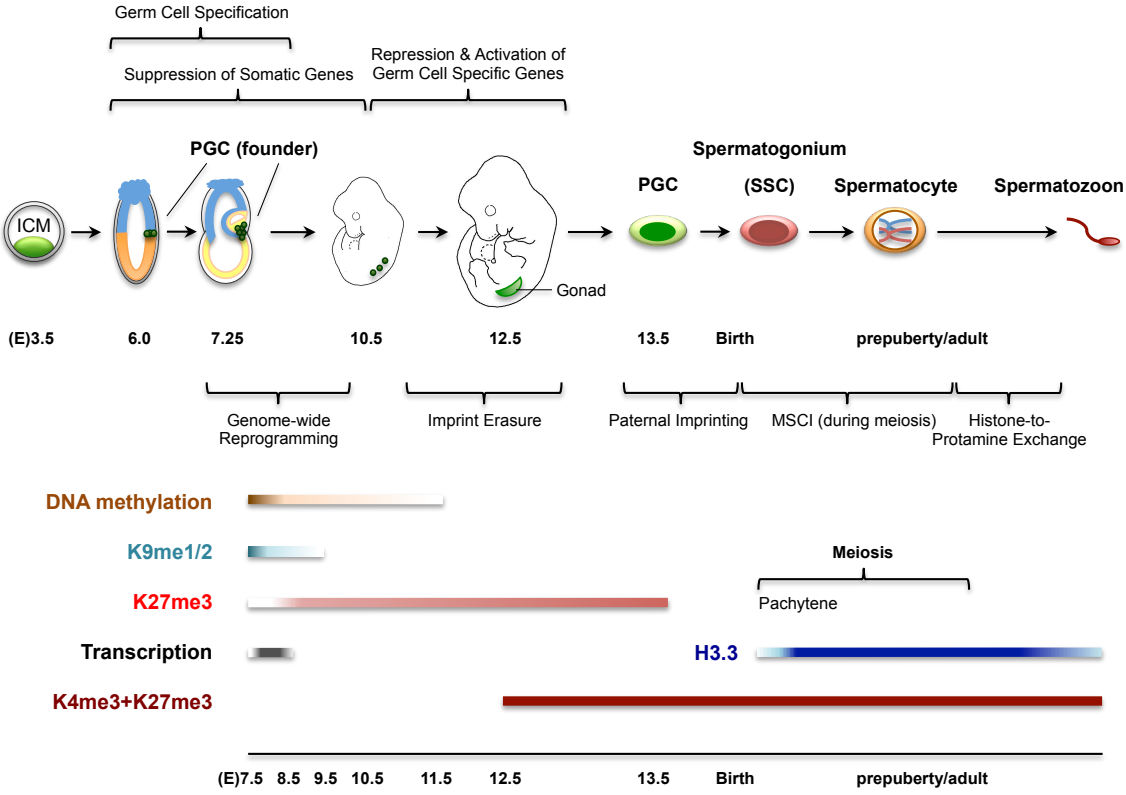


Figure 11. Epigenetic reprogramming during germ cell development in male mouse. Top, the major epigenetic events are listed chronologically. Bottom, Changes in epigenetic modifications or histone variant H3.3. The histone modification or protein level is labeled with light color for low level or dark color for high level, respectively. (Adapted and modified from Figure 1, 2 (Sasaki and Matsui, 2008))

2. Spermatogonial epigenome

Comparing to PGCs, epigenomic study is relatively limit in mammalian spermatogonia in postnatal testes. Immunohistochemical analysis of mouse testes showed that Plzf-expressing spermatogonia contain little H3K27me1 and H3K9me1/2, but H3K27me2/3 and H3K9me3 are detected as punctate foci (Payne and Braun, 2006). Moderate H3K4me3 signal was also detected in A_{undiff} spermatogonia in adult mouse testes (Godmann et al., 2007).

Due to the tiny amount of SSCs in adult mammalian testes, global in-depth analysis of the genetic and epigenetic profiles in SSCs is limited. Recently, Hammoud et al. isolated Thy-1^+ and c-kit^+ adult germline stem cells (AGSCs) from adult mouse testes with a magnetic cell sorting separator. These two cell populations were submitted to in-depth profiling on DNA methylation, histone modifications (H3K4me3 and H3K27me3), and gene expresssion (Hammoud et al., 2014). Interestingly, both Thy-1^+ and c-kit^+ AGSCs share highly similar epigenetic characteristics with ESCs and PGCs (E13.5), including promoter DNA methylation profiles and K4me3+K27me3 bivalent histone modification at many early embryonic gene promoters (Hammoud et al., 2014). This study provides new insights into germ cell biology and rises the interesting notion that developmental pluripotency could be preserved in epigenome and stably transmitted through germline development. However, Thy-1^+ and c-kit^+ AGSCs in this study can hardly represent A_{undiff} and A_{diff} spermatogonia *in vivo*, because both Thy-1 and c-kit are not restrict to spermatogonia but also expressed in many types of somatic cells. As a result, the AGSCs applied in this study were very likely to be contaminated by testicular stroma, and any conclusion derived from this result need to be validated on SSCs with high purity.

Many studies including spermatogonial transplantation show that SSCs are

remarkably stable in long-term culture. Besides of the robust stem cell activity, SSCs retain the euploid karyotype and androgenetic DNA methylation pattern in imprinted genes for at least two years during *in vitro* expansion (Kanatsu-Shinohara et al., 2004; 2005b; Imamura et al., 2006). The stable and robust proliferation makes SSC culture a nice tool for studying spermatogonial epigenome, as unlimit number of cells can be collected for biochemical experiment. However, evaluation of genome-wide epigenetic characteristics in SSC culture has not been reported.

3. Genome-wide epigenetic changes during postnatal spermatogenesis

Germ cells are subject to extensive chromosome reorganization and epigenomic changes during the various stages of spermatogenesis. Genome-wide changes were observed for H3K4me1/2/3 and H3K9me2 at meiotic prophase I, during chromosome pairing, double-strand break (DSB) formation and progression (Kota and Feil, 2010). Several studies indicate that these epigenetic transitions guide meiotic progression. For example, studies in *S. cerevisiae* (Borde et al., 2009), mice, and humans (Baudat et al., 2010; Myers et al., 2010; Parvanov et al., 2010) implicate the role of H3K4me3 in DSB hotspots specification. However, the precise function of these epigenetic marks during meiosis is still not clear (Kota and Feil, 2010).

After the second meiotic division, male mammalian germ cells develop into haploid round spermatids. Genome-wide histone-to-protamine exchange then replace most nucleosomal histones with protamines to generate transcriptionally inert spermatozoa with highly compact nuclei. However, many studies suggested that histone-bearing nucleosomes are retained in about 1–10% of the genome in mouse and human mature spermatozoa (Gatewood et al., 1987; Pittoggi et al., 1999). Moreover,

chromatin immunoprecipitation (ChIP) coupled with high-throughput sequencing revealed that the residual nucleosomes are highly enriched at CpG-rich sequences that lack DNA methylation (Erkek et al., 2013).

Many transcription-associated histone modifications (i.e., H3K4me2/3, H3K27me3, etc.) enrich at loci of embryonic developmental importance in mature spermatozoa genome (Hammoud et al., 2009; Brykczynska et al., 2010). It is suggested that H3K4me3 and H3K27me3 histone modifications poise development genes in spermatozoa as well as in PGCs, pachytene spermatocytes or round spermatids, thus highlighting the maintenance of the bivalent state throughout the germline cycle (Figure 11) (Erkek et al., 2013; Lesch et al., 2013; Hammoud et al., 2014; Lesch and Page, 2014).

Besides of histone modifications, high level of histone variant H3.3 has also been identified in differentiating germ cells in adult mouse testes (Banaszynski et al., 2010). In *Drosophila*, H3.3 expression and deposition is observed in pre-meiotic germ cells, and increase dramatically during the first meiotic prophase when the majority of nuclear H3 is replaced by H3.3 (Akhmanova, AS et al. 1995). H3.3 incorporation is also detected at the XY body during meiotic sex-chromosome inactivation (MSCI) in mouse (van der Heijden et al., 2007). Our immunofluorescence (IF) experiment with H3.3-HA mouse testes also reveals significant H3.3 protein enrichment in post-meiotic round spermatids (Figure 12). Furthermore, recent ChIP-seq result shows that residual nucleosomes in mouse spermatozoa are largely composed of H3.3 and K4me3 modification (Erkek et al., 2013). These findings endow histones and corresponding modifications prime candidacies for epigenetic inheritance across generations.

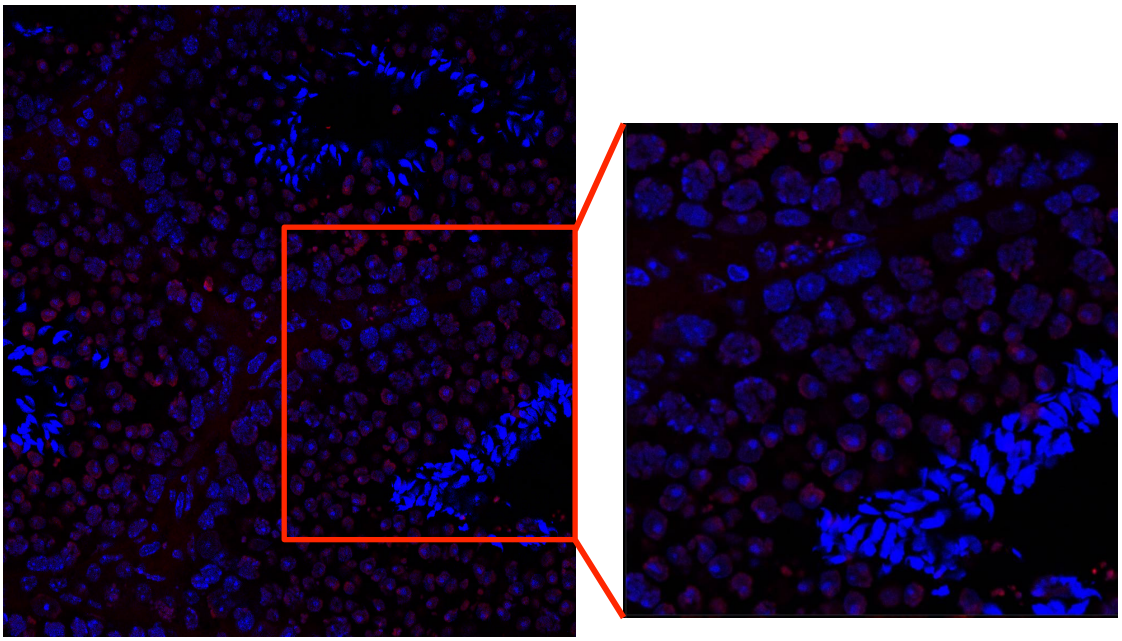


Figure 12. Immunofluorescence with H3.3-HA mouse testes. Red, H3.3-HA. Blue, DAPI.

MATERIALS AND METHODS

I. Experimental Procedures

Mice

The following mouse strains were applied to our study: Pou5f1-GFP JAXR (stock number 004654) (OG) (Szabó et al., 2002), C57Bl6 (B6), C57Bl6/129S (VG) (Seandel et al., 2007), H3.3-HA (Wen et al., 2014c), tetO-4F2A (Carey et al., 2010).

Cell Culture

Mouse SSCs

Primary cultures of mouse SSCs were obtained and cultured as previously described (Seandel et al., 2007). In brief, mice over 3 months-old of the indicated genotypes were euthanized. Seminiferous tubules were collected from detunicated testes and minced on ice. The tissue was enzymatically dissociated with agitation for 30 minutes at 37 °C in a buffer containing 0.017% trypsin (Cellgro), 17mM EDTA (Cellgro), 0.03% collagenase (Sigma-Aldrich), and DNase I (100 μg/ml; Sigma-Aldrich). The cell suspension was then collected and plated in gelatin coated plate in SSC medium containing StemPro-34 (Invitrogen) and supplements as follows: D(+) glucose, 6 mg/ml; BSA, 0.50%; insulin, 25 μg/ml (Sigma-Aldrich); MEM nonessential amino acids, 1× (Gibco); MEM vitamin solution, 1× (Gibco); penicillin (100 U/ml)/streptomycin (100 μg/ml)/amphotericin (0.2 μg/ml) (Media-tech); fetal bovine serum, 1%; L-glutamine, 2 mM (Media-tech); Bovine holo-transferrin, 100 μg/ml (Sigma-Aldrich); β-estradiol, 30 ng/ml (Calbiochem); progesterone, 60 ng/ml (Calbiochem); putrescine, 60 μM (Research Organics); sodium selenite, 30 nM

(Sigma-Aldrich); pyruvic acid, 30 μg/ml (Sigma-Aldrich); d(L)-lactic acid, 1 μg/ml (Baker); β-mercaptoethanol, 50 μM (Gibco); ascorbic acid, 100 μM (EMD); D-biotin, 10 μg/ml (Calbiochem); human GDNF, 10 ng/ml (R&D Systems); human bFGF, 10 ng/ml (Cell Signaling); mouse EGF, 20 ng/ml (R&D Systems). SSC colonies start to appear after about one week. The primary SSCs were then transferred to MEFs (Millipore #PMEF-CF) coated plate and expanded. After at least 10 passages, SSCs were collected by gentle trituration of colonies attached to feeder cells or floating in the medium. To remove residual feeder cells in the collection, SSCs were plated in gelatin-coated plates in fresh SSC medium for 2 hours. All the floating cells were then gently collected, washed once in PBS, and subjected to ChIP experiments or frozen for RNA isolation.

ESCs, MASCs and iPS

To establish MASC or iPS cell lines from corresponding primary reprogrammed cell culture, colonies with typical undifferentiatied ESC morphology were identified by phase microscopy and mechanically separated from the plate using Pasteur pipettes. The colonies were maintained using standard ESCs culture procedures. In brief, cells were all cultured with MEFs in standard ESC medium containing KnockOut DMEM (Gibco) and supplements as follows: KnockOut Serum Replacement, 10% (Gibco); penicillin (100 U/ml)/streptomycin (100 μg/ml)/amphotericin (0.2 μg/ml) (Media-tech); L-glutamine, 2 mM (Media-tech); MEM nonessential amino acids, 1× (Gibco); β-mercaptoethanol, 50 μM (Gibco); ESGRO Leukemia Inhibitory Factor (LIF), 1000 U/ml (Millipore). ESCs, MASCs, or iPS were serially passaged with trypsinization every 2 – 4 days onto fresh inactivated MEFs. All the cells applied for further experiments were collected within 20 passages.

ChIP and antibodies

Chromatin immunoprecipitation (ChIP) was performed as previously described (Goldberg et al., 2010). Briefly, $1x10^7$ cells per experiment were cross-linked for 15min in 1% paraformaldehyde, washed and lysed. Chromatin was sheared using a Bioruptor to create fragments of approximately 150 base pairs, incubated with antibody overnight at 4 °C, then washed and eluted. The eluted chromatin was reverse-cross-linked and column purified. ChIP was performed using the following antibodies: H3K4me3 (Abcam ab8580), H3K27me3 (Abcam ab6002, Cell Signaling #9733), H3K27ac (Abcam ab4729), HA (12CA5 clone prepared by the MSKCC core facility).

Sequence Library Construction and Sequencing

ChIP samples were prepared for sequencing using Illumina TruSeq DNA Sample Preparation Kit according to the standard preparation protocol. RNA samples were prepared for sequencing using Illumina TruSeq RNA Sample Preparation Kit according to the standard preparation protocol (http://www.illumina.com/). Sequencing service was performed on an Illumina Hiseq 2000 sequencer according to the standard Illumina protocol.

ChIP-qPCR

ChIP DNA was diluted 1:100 in H_2O and used for quantitative real-time PCR (qPCR) with 5 μL per reaction. ChIP-qPCR primers used were as follows:

Primer	Sequence (5' -> 3')
Oct4-ChIP-4F	GGAACTGGGTGTGGGGAGGTTGTA
Oct4-ChIP-4R	AGCAGATTAAGGAAGGGCTAGGACGAGAG
Oct4-ChIP-10F	GGAGTCCCCTAGGAAGGCATTAATAGTTT
Oct4-ChIP-10R	GGATTCTCTCGGCTTCAGACAGACTTT
Oct4-ChIP-12F	CCCTTCCAGAGCCCCTTTCAGTAA
Oct4-ChIP-12R	GCACCAGGGTCTCCGATTTGCATA
Oct4-ChIP-15F	AGAGTTGGGAGCAGCCATGTAGGTT
Oct4-ChIP-15R	TGGCCTGTGGTGGTACATGACCTTA
Ch8-ChIP-F	AAGGGGCCTCTGCTTAAAAA
Ch8-ChIP-R	AGAGCTCCATGGCAGGTAGA
Gata6-ChIP-F	CCTTCCCATACACCACAACC
Gata6-ChIP-R	CCCCTCCTTCCAAATTAAGC
Actin-ChIP-3F	TTGATAGTTCGCCATGGATGACGA
Actin-ChIP-3R	ATCGATCCCCAAGAAAACCCCA
Nanog-ChIP-4F	TCTGCTGCCTAAGCTCTTGTGCTGT
Nanog-ChIP-4R	CTGCTCTTTCCAATCCCTTCTCCC
HoxA1-ChIP-F	TCACTGAGTGATTGGATCCTGC
HoxA1-ChIP-R	GGAGGAAGTGAGAAAGTTGGCAC
HoxA2-ChIP-F	GACAAGGTTGAAATTGGACCG
HoxA2-ChIP-R	CAAATTGTCATTGGGCAGAAGC
HoxA4-ChIP-F	CTCTGGAATAAAACGAAGGAGGC
HoxA4-ChIP-R	GGACAAAGAATCAAAGGGCGAG
HoxA6-ChIP-F	CTTTCCTTTTTTGCCTTCATGG
HoxA6-ChIP-R	TTGTCAGGTTTCCTGTTTGGG
HoxA7-ChIP-F	AACCCTTCCCCTAAACGCCTC
HoxA7-ChIP-R	AAAAGGTCGCCAGTCTTCCAG
HoxA9-ChIP-F	ATCTGTATGCCTAGTCCCGCTCC
HoxA9-ChIP-R	TTGATGTTGACTGGCGATTTTC
HoxA10-ChIP-F	CTGGCTCTTGAACCTGTACCCC
HoxA10-ChIP-R	CAAGGGTGCTTCCAAATAGTC
HoxA11-ChIP-F	GGAAGCAACAGATCGTCACTCG
HoxA11-ChIP-R	TGAGTTACACCGGCGATTACG
HoxA13-ChIP-F	CTATGACAGCCTCCGTGCTC
HoxA13-ChIP-R	CCCCTTCCATGTTCTTGTTG

II. Data Analytic Methods

RNA-seq Data Processing and Analysis

Reads from RNA-seq were aligned to mouse genome version mm9 using TopHat (Trapnell et al., 2009), and fragments per kilobase of transcript per million fragments mapped (FPKM) were identified using Cufflinks with upper-quartile and GC-normalization (Trapnell et al., 2010). Duplicate reads and reads aligning to more than one location were excluded. Gene expression was reported as log2 transformed FPKM value.

Analyses applied in chapter I

Genes with multiple isoforms and TSS were summarized for FPKM by common transcription start site (TSS) to get a total of 23,183 unique transcripts. Only genes within 2 (log2) fold change between two biological replicates of MASCs were considered to have consistent expression after reprograming; there were averaged based on expression values, and applied for subsequent analyses. Using this gene set, we selected all genes (3,316 unique transcripts) with expression differences of at least 2 (log2) fold between any pair of the cell types, SSCs, MASCs, and ESCs. These gene expression values were used as input for hierarchical clustering (centered on gene, uncentered on cell type, centroid linkage) by Cluster 3.0 (de Hoon et al., 2004).

Analyses applied in chapter II

Genes with multiple isoforms and TSS were summarized for FPKM by common TSS to get a total of 24,033 unique transcripts. For genes with H3.3 replacement in exons, expression at each time point after induction was compared with that at day 0, and increase over 1 (log2) fold was considered as increased expression. Genes *de novo* activated in iPS were selected by over 2 (log2) fold increase in iPS than in MEFs (day

0), and expression in MEFs (day 0) is below 5 (log2 transformed FPKM value). There were 2,069 genes selected with these criteria as *iPS activated genes*. Genes remaining stable low expression in both MEFs and iPS were selected by within 1 (log2) fold change between MEFs (day 0) and iPS, and expression in MEFs (day 0) is below 5 (log2 transformed FPKM value). Within 10,112 unique transcripts matching the criteria, 2,000 transcripts were randomly selected as *iPS stable genes* and subjected to further analyses. Expression changed genes during 10 days of reprogramming were selected as any gene with over 2 (log2) fold expression difference between any two time points from day 0 to day 10. All the rest were considered as stable expression genes during 10 days of reprogramming.

ChIP-seq Data Processing and Analysis

ChIP-seq reads were aligned to the reference mouse genome (mm9, NCBI Build 37) using the BWA program (Li and Durbin, 2009), and PCR duplicates were removed by Picard (http://picard.sourceforge.net/). Unique reads that mapped to a single best-matching location with no more than 4% of the read length of mismatches were kept and used to study genome-wide enrichment of specific histone modification. Sequence data was visualized with IGV by normalizing to 1 million reads (Thorvaldsdóttir et al., 2013).

Analyses applied in chapter I

The software ChIPseeqer 2.0 was applied to the ChIP-Seq data with sequencing data from input DNA as control to identify genomic enrichment (peak) of specific histone modifications (Giannopoulou and Elemento, 2011). Promoters were defined as +/- 2 kb from transcription start site (TSS). The promoter chromatin state was determined by overlap with significant H3K4me3 and/or H3K27me3 peaks (t=5, with t as the

significance negative log p-value [ratio] threshold for peaks) (FDR<0.05). Promoter ChIP-seq Read Intensity ratio for histone Modification K4me3 and K27me3 (PRIMs) was calculated as $PRIM = log2\frac{K4me3+a}{K27me3+a}$, with K4me3 and K27me3 as average read counts from H3K4me3 and H3K27me3 sequences at the same promoter region and constant 'a' equals to 0.001. Only promoters with detectable K4me3 or K27me3 peaks were evaluated by PRIMs.

An enhancer region was defined as any genomic locus with H3K27ac enrichment but no H3K4me3 enrichment (K27ac+K4me3-). To identify enhancer loci, K4me3 peaks were extended 1 kb each way. All H3K27ac peaks not overlapping with extended H3K4me3 peaks, known gene body, transcription start (TSS) and end site (TES) were selected. Selected H3K27ac peaks within a 500 bp interval were merged together as enhancers. Enhancer regions from SSCs (OG), MASCs (OG), and ESCs (CMP10) were merged together. Each enhancer was annotated with all transcripts within a distance of 100 kb from TSS or TES. Any enhancer overlap with K27ac peaks in a given cell type was considered to be active in that cell type.

Analyses applied in chapter II

To study global H3.3 enrichment, we segmented the whole mouse genome into non-overlapping 1kb windows, and calculated H3.3B-HA sequence read count per window with SNPseeqer developed in Dr. Elemento's lab. Results from all selected cell types were compared by correlation clustering. To identify global H3.3 replacement sites during MEFs reprogramming, results between any time point after doxycycline induction and day 0 were compared, and genomic loci with over 2 folds read count increase (normalized by total read count of each sample) were selected as H3.3 replacement sites. All H3.3 replacement sites were annotated to genome, with

genomic regions defined as: promoter, -2kb ~ TSS; TES, -2kb ~ 2kb around TES; distal region, over 2kb from transcript.

To study H3.3B-HA and H3K27me3 enrichment at promoter and enhancer regions, we applied ChIPseeqer 2.0 developed in Dr. Elemento's lab to calculate average sequence read count at selected loci, and reported as log2 transformed value (Giannopoulou and Elemento, 2011). Promoter was defined as -2 kb to transcription start site (TSS) of each transcript. Genomic loci with average read count at any time point after doxycycline induction higher than day 0 over certain cut-off (H3.3B-HA, 0.15; H3K27me3, 0.2) are considered as increased with ChIP signals. Genomic loci with average H3K27me3 read count at day 0 higher than that at any time point after doxycycline induction over 0.2 are considered as decreased with H3K27me3 modification. Genomic loci underwent both increase and decrease of H3K27me3 during 10 days of reprogramming are considered as decreased with H3K27me3 modification. In MEFs (day 0), promoter with low or high H3K27me3 is selected by below or above 0.2.

To study promoters with 'bivalent' modification in ESCs, we applied ChIPseeqer 2.0 to identify significant peaks (t=10) on H3K4me3 and H3K27me3 ChIP-seq dataset. The peaks were then annotated to promoter, which is defined as -2kb ~ 2kb around TSS as previousely described (Banaszynski et al., 2013). Promoters with both significant H3K4me3 and H3K27me3 peaks are considered as ESCs 'bivalent' promoters.

To study enhancer regions potentially function in gene activation during reprogramming, we identified H3K27ac modified active enhancers in both MEFs and ESCs (adapted from Creyghton MP, et al., 2010). Significant H3K27ac peaks (t=5) were identified by ChIPseeqer 2.0, peaks from both MEFs and ESCs and within 500

bp distance were merged together, those located at least 2 kb away (edge to edge) from TSS and outside open reading frame were collected as enhancers potentially function in reprogramming. Enhancers within 100 kb distance from TSS or TES of a transcript were annotated to that gene.

Gene Ontology Analysis

Gene Ontology (GO) analysis was done using iPAGE with mouse GO annotation database (Goodarzi et al., 2009). 31,881 GO terms were examined. Only non-electronic annotations were used in iPAGE and only categories with less than 300 genes were analyzed. IPAGE uses randomized simulations to ensure FDR<5%.

Motif Analysis

Enrichment of known motifs within promoter and enhancer regions was analyzed with HOMER with default parameters and a fragment size of 200 bp. All known motifs used in our study were defined by HOMER (Heinz et al., 2010).

Statistical Analysis

Statistical Analysis was performed in R (version 3.1.0) statistical framework (R Development Core Team, 2008). R packages applied for analysis and graph include limma (3.20.8), gplots (2.14.1), ggplot2 (1.0.0), VennDiagram (1.6.7).

CHAPTER ONE

DYNAMIC AND STABLE EPIGENETIC PROFILES SIGNIFY CELL FATE CONVERSION IN UNIPOTENT MOUSE SPERMATOGONIAL STEM AND PROGENITOR CELLS

I. Summary

Mammalian spermatogonial stem and progenitor cells (SSCs) are precursors of all subsequent germ cells in the adult male gonad. During *in vitro* expansion, these unipotent adult stem cells can spontaneously convert to multipotent adult spermatogonial-derived stem cells (MASCs) without ectopic expression of any transcription factors. The underlying mechanisms of this reprogramming event are still poorly understood. Here, We report an integrative analysis of both the transcriptomes and epigenomes of mouse SSCs and MASCs. We find in SSCs that many genes essential to maintenance and differentiation of embryonic stem cells (ESCs) are enriched with both histone H3 lysine 4 and lysine 27 trimethylation modifications (K4me3+K27me3) at promoter regions. Despite significantly increased expression after reprogramming, most of these genes maintain 'bivalent' promoter modification, with only a few ESCs pluripotency gene promoters selectively depleted of the K27me3 modification in MASCs. Quantitative epigenomic analysis further suggests that promoter histone modifications are restrictively changed at signature genes for ESCs and germ cells but remain repressive for most somatic genes, bestowing MASCs

with ESC-like promoter chromatin states. At enhancer regions, the core pluripotency circuitry is activated partially in SSCs and completely in MASCs, concomitant with global erasure of germ cell-specific enhancer activity and initiation of an embryonic-like program. Furthermore, long-term *in vitro* cultured SSCs maintain epigenomic characteristics with germ cells *in vivo*. Our observations suggest that unipotent SSCs encode their innate developmental flexibility by means of the epigenome and that both the promoter chromatin state, and the activity of cell-type-specific enhancers is a prominent feature of SSC reprogramming.

II. Introduction

As the precursors of germ cells, mammalian spermatogonial stem and progenitor cells (SSCs) undergo unipotent differentiation in the adult male gonad, while still possessing the ultimate developmental potency to propagate across generations. During *in vitro* expansion, mouse SSCs, despite being unipotent, are uniquely capable of abrogating lineage commitment and spontaneously conversion to multipotent adult spermatogonial-derived stem cells (MASCs), which share many common features with pluripotent embryonic stem cells (ESCs), including the capacity to induce teratomas and contribute to chimeric animals (Figure 1A) (Kanatsu-Shinohara et al., 2004; Li et al., 2007; Seandel et al., 2007). To date, this is the only known reprogramming event that converts unipotent adult stem cells back to a near pluripotent state without delivery of exogenous genes or gene products, which distinguishes it from formation of induced pluripotent stem (iPS) cells (Bernstein et al., 2006; Takahashi and Yamanaka, 2006; Mikkelsen et al., 2007; Takahashi et al., 2007). These observations indicate that intrinsic genetic and epigenetic features rather than enforced expression of ectopic transcription factor(s) are responsible for

reprogramming of SSCs. However, SSC conversion into MASCs is a rare event, and the underlying mechanisms remain largely unknown.

One possible explanation for the spontaneous loss of lineage commitment is that SSCs may preserve a latent ESC-like gene expression program. Indeed, upon germline specification in the mouse embryo, somatic genes are mainly repressed in primordial germ cells (PGCs), while expression of several ESC signature transcription factors (e.g., Pou5f1/Oct4, Klf4, Esrrb) is reactivated and preserved at modest levels in SSCs in the adult testis (Pesce et al., 1998; Jaenisch and Young, 2008). Pou5f1/Oct4, together with Sox2 and Nanog, form the core pluripotency circuitry in ESCs to sustain stem cell self-renewal and control the expression of many differentiation genes (Boyer et al., 2005; Loh et al., 2006; Creyghton et al., 2010). As the precursors of all subsequent germ cells, SSCs also express spermatogenesis-specific genes (e.g., Zbtb16/Plzf, Piwil4) but repress regulators of somatic cell development (e.g., Prmt8) (Costoya et al., 2004; Carmell et al., 2007).

Among other mechanisms that drive spontaneous conversion of SSCs into MASCs, the histone modification state plays a key role in transcriptional regulation and cell type specification (Goldberg et al., 2007; Surani et al., 2007). Extensive studies have shown that histone modifications are closely associated with transcriptional gene regulation. Notably, trimethylation on histone H3 lysine 4 (K4me3) at promoters and acetylation on histone H3 lysine 27 (K27ac) at enhancers, respectively, correlate with gene expression, while Polycomb Repressive Complex 2 (PRC2)-mediated trimethylation on histone H3 lysine 27 (K27me3) correlates with gene silencing (Li et al., 2007). In ESCs, K4me3 and K27me3 co-localize at promoters of many developmental genes that control stem cell differentiation to all somatic and germline lineages (Bernstein et al., 2006; Mikkelsen et al., 2007). Therefore, the

K4me3+K27me3 'bivalent' modification is suggested to poise promoters in a repressive state, while the switch to a K4me3 or K27me3 univalent modification state is generally believed to direct gene activation or complete silencing, respectively, during stem cell differentiation (Jaenisch and Young, 2008). Enhancers also confer *cis*-regulation to gene expression by recruiting specific transcription factors, and enhancer activity is tightly controlled in a cell-type-specific manner (Creyghton et al., 2010). The presence of the K27ac modification at genomic regions without the K4me3 modification is considered to denote active enhancers, although it has also been reported that strong enhancers are marked by both K27ac and K4me3 modifications.

Based on these observations, we hypothesize that the epigenetic landscape of SSCs is plastic and under certain as yet unrecognized conditions may cause conversion back to a ESC-like state. We performed transcriptome sequencing (RNA-seq) and found that MASCs are distinguished from SSCs by reactivation of somatic lineage-specific genes and silencing of spermatogenesis regulators. To elucidate the underlying molecular mechanisms, we performed chromatin immunoprecipitation and massively parallel sequencing (ChIP-Seq) on SSCs and MASCs to identify changes in histone modifications at both promoters and enhancers of reprogrammed genes. These genome-wide studies revealed that K4me3 and K27me3 co-localize to a significant number of developmental gene promoters in SSCs, and over two-thirds of these promoters are bivalent modified in both MASCs and ESCs. Notably, in MASCs, which function like pluripotent stem cells (e.g., in blastocyst chimerism), K27me3 is lost from genes crucial to early embryogenesis and stem cell maintenance but is increased at promoters of most germ cell-specific genes with potential gene-specific recruitment of the PRC2 (Suz12) complex. This dynamic epigenetic alteration closely correlates with gene expression changes and endows MASCs with ESC-like promoter chromatin states. At enhancer loci, we noted that the germ cell epigenetic "signature"

is largely lost in MASCs compared to SSCs, while ESC-specific enhancers are partially activated after reprogramming. Intriguingly, unipotent SSCs share substantial enhancer activity with multipotent MASCs and pluripotent ESCs. Indeed, active enhancers shared by SSCs and MASCs are predicted targets of many ESCs signature transcriptional regulators, including core pluripotency regulators such as Pou5f1/Oct4, Sox2, and Nanog. These transcription factors are also involved in *de novo* enhancer activation in MASCs.

Taken together, our study provides global histone modification maps of mouse SSCs and MASCs, and reveals dynamic changes in epigenetic signatures before and after SSC reprogramming. These results offer new insight into transcription factor-independent epigenetic regulation during mammalian cell reprogramming from a unipotent to a multipotent state and suggest several strategies to improve SSC reprogramming efficiency.

III. Result

1. SSCs, MASCs and ESCs share similar epigenomes

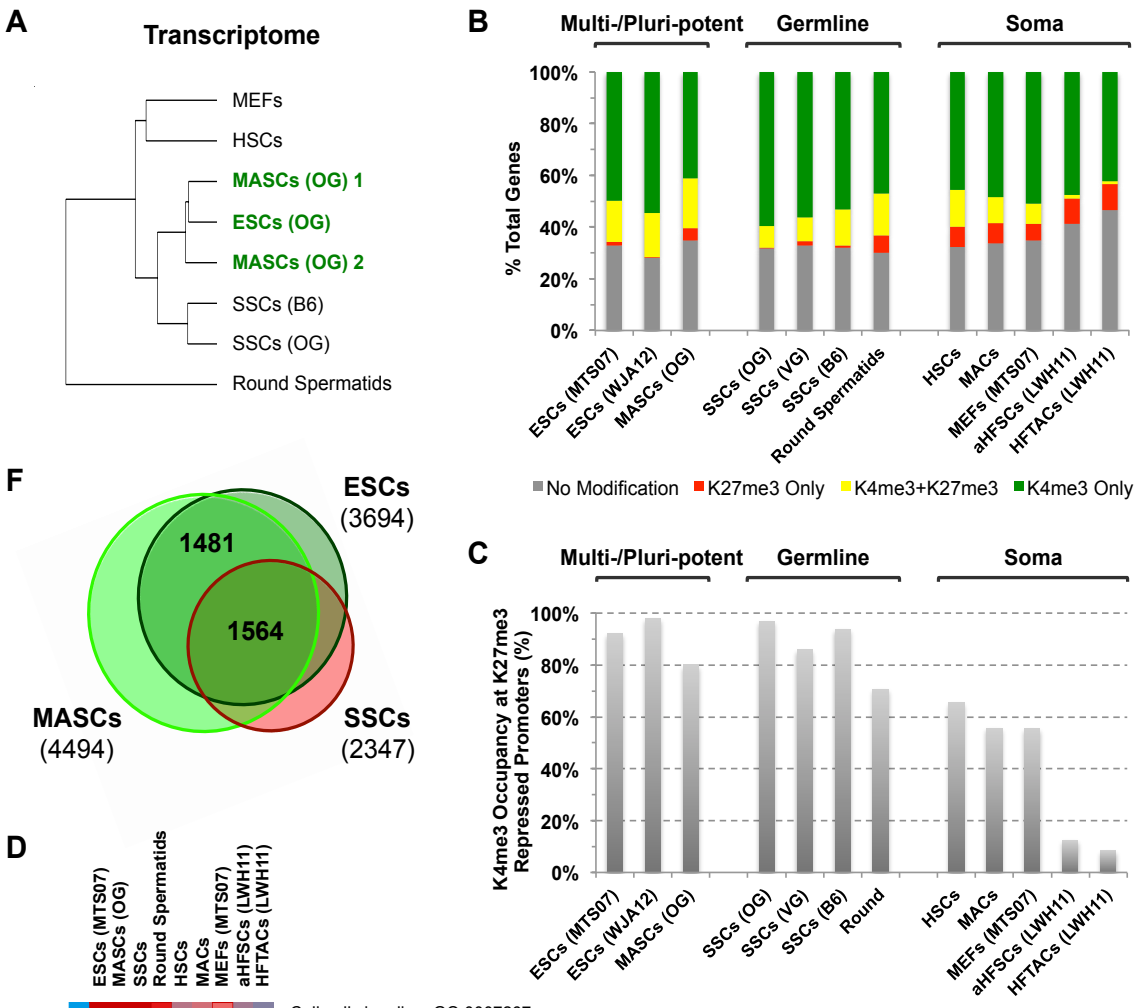
We first compared global gene expression among SSCs, MASCs, and ESCs using hierarchical clustering of RNA-seq profiles collected from each of these cell types (Table 1.1). To ensure consistent transcriptional profiling, we selected two MASCs clones that efficiently form tri-lineage teratomas, a key criteria for pluripotency. These clones were isolated for RNA-seq. Consistent with previous studies, MASCs were transcriptionally similar to ESCs (Kanatsu-Shinohara et al., 2004; Seandel et al., 2007). Moreover, in comparison with somatic multipotent stem cells (hematopoietic stem cells, HSCs), differentiated germ (round spermatids) or somatic (mouse

embryonic fibroblasts, MEFs) cells, SSCs were closely associated with MASCs and ESCs, irrespective of genetic background (Figure 1.1A).

To investigate how the transcriptomes and epigenomes were coordinated for reprogramming of SSCs, we performed ChIP-Seq for histone modifications that are closely associated with transcriptional activation (K4me3, K27ac) and repression (K27me3) (Table 1.2). As expected, peak detection with ChIPseeqer-2.0 (FDR<0.05) (Giannopoulou and Elemento, 2011) revealed that both MASCs and ESCs were enriched with K4me3+K27me3 bivalent histone modifications at promoters of many developmental genes, an epigenetic signature of stem cell pluripotency (Figure 1.1B – F, Table 1.3, 1.4) (Mikkelsen et al., 2007). Notably, over 80% of the K27me3-marked promoters also possessed K4me3 modifications in SSCs (Figure 1.1C), and two-thirds (1,564/2,347) of these genes bearing bivalent modifications in SSCs were shared with MASCs and ESCs (Table 1.5); such genes regulate embryonic development into both germline and somatic lineages (Figure 1.1E, F).

Figure 1.1. SSCs, MASCs and ESCs share similar transcriptomes and promoter bivalency.

(A) Dendrogram based on expression of all 23,183 protein-coding and noncoding genes. Green, multipotent or pluripotent stem cells.
(B) Promoter categories basing on peak detection result from K4me3 and K27me3 modifications. Green, K4me3 peak only. Red, K27me3 peak only. Yellow, both K4me3 and K27me3 peaks detected. Grey, no K4me3 or K27me3 peaks detected.
(C) Percentage of K4me3+K27me3 bivalent promoters in K27me3 marked promoters.
(D) GO enrichment in bivalent genes identified from different cell types. Red, over representation. Blue, under representation.
(E) GO enrichment in bivalent genes shared or unique in SSCs, MASCs, and ESCs. +, promoter bivalency detected in certain cell type.
(F) Overlap of bivalent promoters identified by peak detection in different cell types.


A
Transcriptome
MEFs
HSCs
MASCs (OG) 1
ESCs (OG)
MASCs (OG) 2
SSCs (B6)
SSCs (OG)
Round Spermatids
B
Multi-/Pluri-potent
Germline
Soma
% Total Genes
100%
80%
60%
40%
20%
0%
ESCs (MTS07)
ESCs (WJA12)
MASCs (OG)
SSCs (OG)
SSCs (VG)
SSCs (B6)
Round Spermatids
HSCs
MACs
MEFs (MTS07)
aHFSCs (LWH11)
HFTACs (LWH11)
No Modification
K27me3 Only
K4me3+K27me3
K4me3 Only
C
Multi-/Pluri-potent
Germline
Soma
K4me3 Occupancy at K27me3 Repressed Promoters (%)
100%
80%
60%
40%
20%
0%
ESCs (MTS07)
ESCs (WJA12)
MASCs (OG)
SSCs (OG)
SSCs (VG)
SSCs (B6)
Round
HSCs
MACs
MEFs (MTS07)
aHFSCs (LWH11)
HFTACs (LWH11)
F
ESCs
(3694)
1481
1564
MASCs
(4494)
SSCs
(2347)
D
ESCs (MTS07)
MASCs (OG)
SSCs
Round Spermatids
HSCs
MACs
MEFs (MTS07)
aHFSCs (LWH11)
HFTACs (LWH11)
Cell-cell signaling, GO:0007267
Anterior/posterior pattern formation, GO:0009952
Cell morphogenesis involved in differentiation, GO:0000904
Morphogenesis of an epithelium, GO:0002009
Dorsal/ventral pattern formation, GO:0009953
Neuron fate commitment, GO:0048663
Mesenchyme development, GO:0060485
Development of primary sexual characteristics, GO:0045137
Embryonic skeletal system development, GO:0048706
E
SSCs + - + + + - -
MASCs + + - - + + -
ESCs + + + - - - +
Depletion
Enrichment
Ectoderm development, GO:0007398
Regulation of mesenchymal cell proliferation, GO:0010464
Organ formation, GO:0048645
Embryonic placenta development, GO:0001892
Formation of primary germ layer, GO:0001704
Endoderm development, GO:0007492
Pancreatic B cell differentiation, GO:0003309
in utero embryonic development, GO:0001701
Morphogenesis of an epithelium, GO:0002009
Enzyme linked receptor protein signaling pathway, GO:0007167
Reproductive structure development, GO:0048608
Regulation of nervous system development, GO:0051960
Positive regulation of cell differentiation, GO:0045597

Notably, a significant number of bivalent genes in SSCs was transcriptionally activated, but not repressed, in MASCs (Figure 1.2, Table 1.5). This result suggests both that unipotent SSCs possess a plastic chromatin configuration characterized by bivalent histone modifications, a feature conserved with other cell types in a pluripotent developmental state, and that bivalent genes in SSCs are generally mainly poised for activation before converting to MASCs.

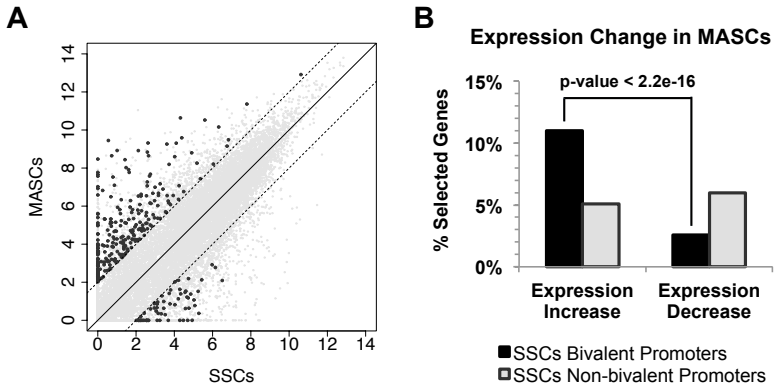


Figure 1.2. SSC bivalent genes are mainly activated in MASCs.

(A) Comparison of global gene expression profiles between SSCs and MASCs. Black dots, peak detection identified SSC bivalent genes.

(B) Percentage of genes with expression increase or decrease in MASCs. Black bar, peak detection identified SSC bivalent genes. Grey bar, genes modified with either K4me3, K27me3, or none of the two modifications in SSCs.

To evaluate the relative abundance of these two histone modifications at the same genomic loci, we next quantified the ratio between K4me3 and K27me3 modifications at promoters with either or both peaks, referred to as the Promoter ChIP-seq Read Intensity ratio for histone Modification (PRIM) (see Methods). A PRIM cutoff that effectively distinguished K27me3-repressed promoters ($PRIM^{low}$, either K4me3+K27me3 or univalent K27me3) from active promoters ($PRIM^{high}$, univalent

K4me3) was also identified for each cell type (Figure 1.3). Cluster analysis using PRIM values showed that MASCs and ESCs were close to each other in terms of K4me3- and K27me3-defined promoter chromatin states, and unipotent SSCs were closer to pluripotent cells than other cell types in this study (Figure 1.4).

Thus, these results demonstrate that MASCs closely resemble ESCs with respect not only to global gene expression but also promoter histone modification and relative K4me3/K27me3 enrichments. Moreover, a significant pluripotency signature in the form of bivalent promoter modification is embedded in SSCs, the precursors of MASCs. By contrast, this pluripotency-associated epigenetic signature was less prevalent in other adult stem and differentiated cells that had been studied (Figure 1.1B – D).

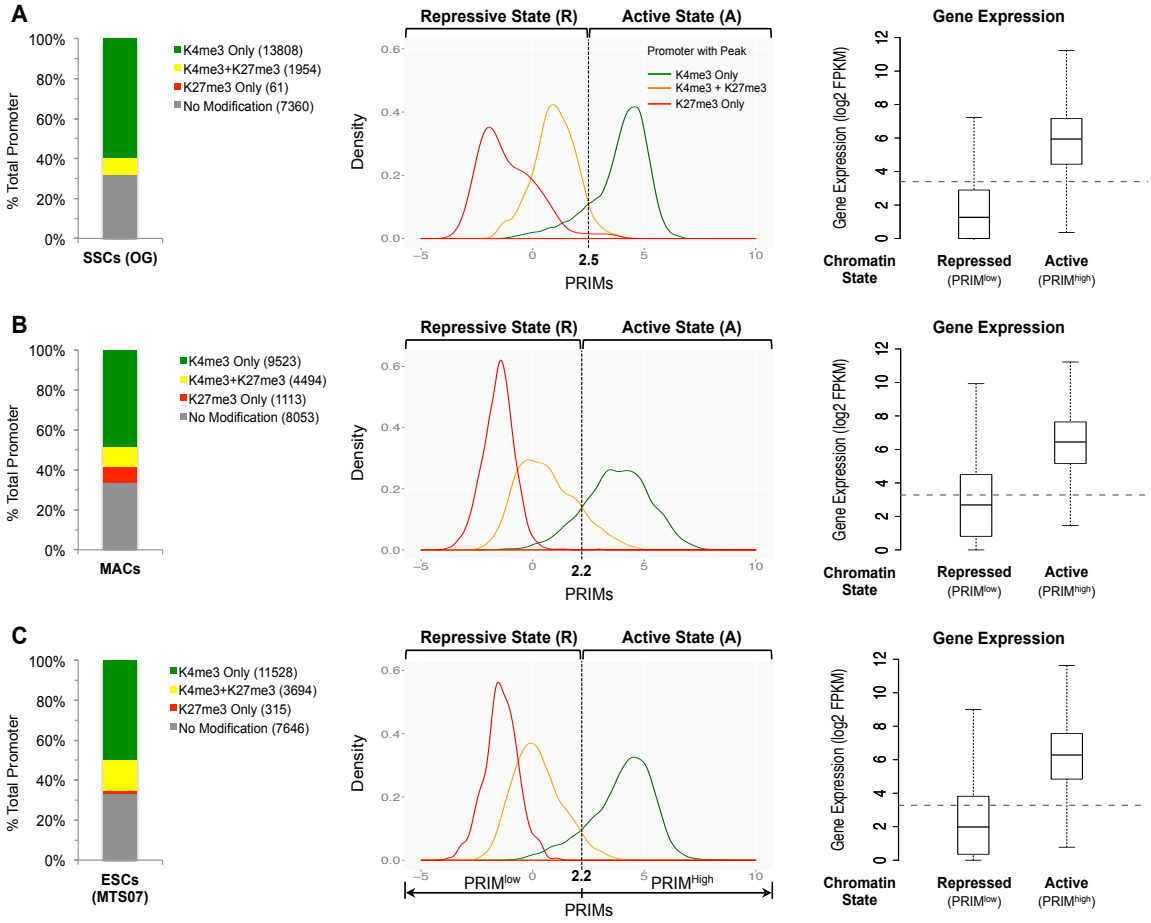


Figure 1.3. Distinguish active and repressed promoter chromatin states by PRIMs in (A) SSCs, (B) MASCs, and (C) ESCs.

For each cell type, all promoters were assigned to each of the four chromatin state categories basing on peak detection result as in the bar graph (Left). Probability density distribution of PRIMs was made for each of the three states, K4me3 only, K4me3+K27me3, and K27me3 only (Middle). The cut-off value between K4me3 and K4me3+K27me3 or K27me3 groups was selected by the crossover between K4me3 and K4me3+K27me3 plot. Y-axis, probability density. X-axis, PRIMs. Cut-off PRIM value, SSCs=2.5, MASCs=2.2, ESCs=2.2. Expression of genes in active (> PRIM cut-off) and repressed (<= PRIM cut-off) promoter chromatin states is shown in boxplot (Right). Dashed line, median of all gene expression value.

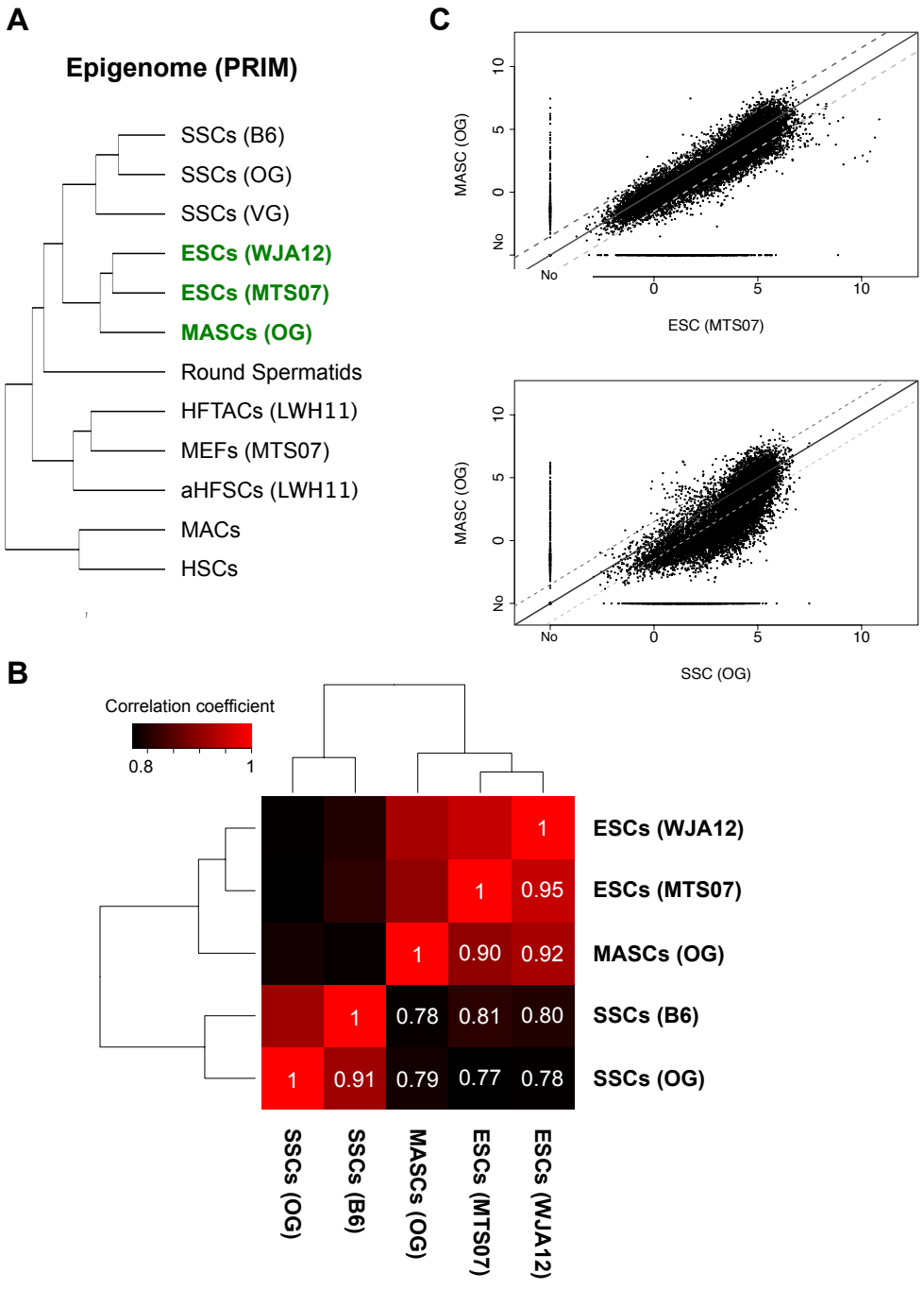


Figure 1.4. Promoter histone modification correlation analysis with PRIMs.

(A) Dendrogram based on PRIMs of all protein-coding/noncoding gene promoters.
(B) Correlation of promoter PRIMs for all ChIP-seq datasets. Only promoters with detectable K4me3 or K27me3 were calculated for PRIMs. Correlation heat map was built basing on Pearson's correlation coefficients between every pair of data.
(C) Plot on PRIMs between MASCs and ESCs (up) or MASCs and SSCs (down) for all promoters.

2. SSC reprogramming involves activation of early embryonic genes and silencing of spermatogenesis-specific genes

To elucidate the mechanisms underlying spontaneous reprogramming of SSCs, we then focused on 3,316 genes differentially expressed (>2-fold change) among SSCs, MASCs, and ESCs (Table 1.6). Hierarchical clustering assigned them into six classes as described below (Figure 1.5A).

Over 90% (2,999/3,316) of the selected genes shared expression between MASCs and ESCs but differed from SSCs and were therefore considered as "reprogramming complete" genes (Figure 1.5A). Gene Ontology (GO) analysis showed that genes activated after reprogramming were mainly involved in ESC self-renewal and early embryonic development toward both somatic and germline lineages (e.g., Pou5f1/Oct4, Nanog, Fgf4, Class I) (Figure 1.5B, Table 1.6). Conversely, genes repressed in reprogrammed MASCs were enriched with regulators of germ cell differentiation and meiosis (e.g., Zbtb16/Plzf, Piwil4, Class II) (Figure 1.5B, Table 1.6).

The remaining 10% of the selected genes were not equally expressed in MASCs and ESCs. Among them, 211 "reprogramming incomplete" genes maintained similar expression between SSCs and MASCs but were relatively overexpressed (Class III) or repressed (Class IV) in ESCs (Figure 1.5A). Notably, a few transcription factors that function in mesoderm differentiation (e.g., Snai1, Ets1, Class IV) exhibited higher expression in MASCs but were repressed in ESCs (Figure 1.5B, Table 1.6). We also found a few genes uniquely overexpressed (Class V) or repressed (Class VI) in MASCs. In particular, Class V genes were significantly enriched with epithelial lineage regulators (e.g., Gata6, Wnt4) (Figure 1.5B, Table 1.6). Overexpression of these mesenchymal and epithelial signature transcription factors in MASCs could be

due to cell differentiation after long-term passage in culture and might bias the relative contribution to different germ layers during teratoma formation. Indeed, mesenchymal cells frequently differentiated from MASCs in my lab (data not shown), as reported also by other groups (Baba et al., 2007).

In summary, SSC reprogramming involved induction of early embryonic genes that function in ESC self-renewal and differentiation to both somatic and germline lineages, while expression of spermatogenesis-specific genes was reduced in MASCs.

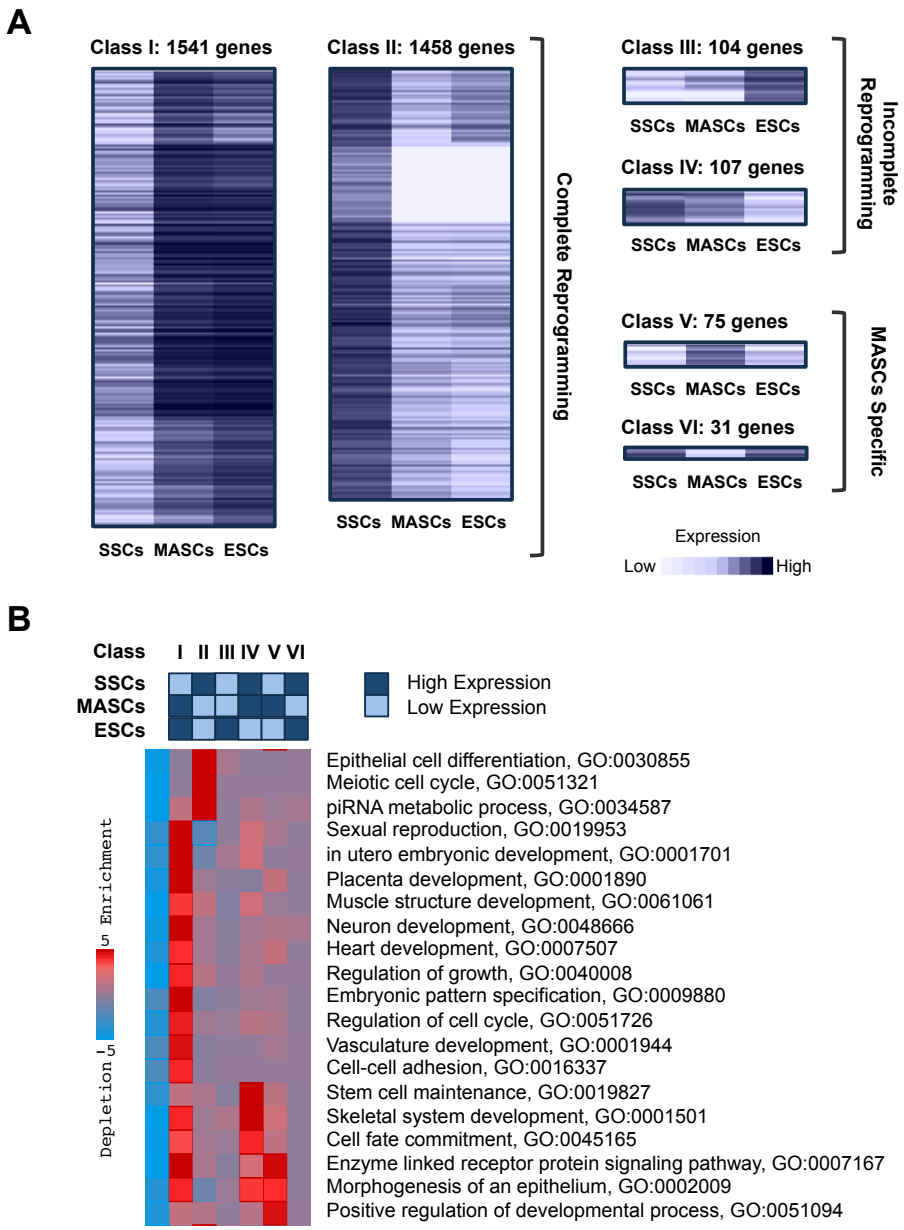


Figure 1.5. Differentially expressed genes among SSCs, MASCs, and ESCs.

(A) Differential gene expression profiling among ESCs, MASCs, and SSCs. Genes with over two-fold (log2) difference between any pair of samples were selected and subjected to hierarchical clustering. Blue and white indicate relative higher and lower expression, respectively.

(B) Gene ontology (GO) enrichment using iPAGE. Top: genes classified as in Figure 1.5A. Dark blue, high expression. Light blue, low expression. Bottom, enrichment of GO biological functions in each gene class. Enrichment and depletion of certain GO functions (shown on the right) are measured by hypergeometric p-value (log10-transformed). The first column contains control genes that do not belong to any of the other 6 classes. Red and blue indicate relative over and under representation, respectively.

3. Bivalent promoter modification poises activation of somatic and ESC signature genes (Class I) in SSCs, with only a few switching to active chromatin states in MASCs

We hypothesized that reprogramming of SSCs without enforced expression of transcription factors would be associated with concomitant chromatin changes. Therefore, we first investigated the promoter chromatin states of genes that are completely activated in MASCs (Class I) (Figure 1.5A, 1.6A). Strikingly, Class I gene promoters largely shared similar K4me3- and/or K27me3-marked chromatin states among SSCs, MASCs, and ESCs (Pearson's correlation=0.81) (Figure 1.6B – D). Despite significantly increased expression after SSC reprogramming, there were 22.5% promoters possessed both K4me3 and K27me3 modifications in SSCs and MASCs (347 promoters, $MASC^{Stable\ I}$), including most of the bivalent promoters found in SSCs by peak detection (black dots, Figure 1.6D, 1.7A, Table 1.6). These $MASC^{Stable\ I}$ genes mainly function in stem cell differentiation and embryonic organ development into somatic lineages (e.g., Esrrb, Prmt8) (Figure 1.6A, E, 1.7C). Conversely, only a small subset of Class I gene promoters acquired *de novo* modifications, K4me3 in particular, in MASCs (63 promoters, $MASC^{M}$), or changed from repressive (SSCs) to active (MASCs) chromatin state with complete erasure of K27me3 (108 promoters, $MASC^{Active}$) (Figure 1.6A, D, 1.7A, Table 1.6). $MASC^{M}$ and $MASC^{Active}$ subsets included many ESC signature genes associated with stem cell identity, e.g., Nanog ($MASC^{M}$) and Zic3 ($MASC^{Active}$) (Figure 1.6A, E, 1.7C). Compared to $MASC^{Stable\ I}$ genes, $MASC^{M}$ and $MASC^{Active}$ genes were highly expressed in MASCs and ESCs, indicating a strong impact of chromatin state changes on transcriptional regulation (Figure 1.6A, 1.7B). These three types of gene clusters included the majority of the pluripotency and developmental regulators activated in MASCs (Figure 1.7C).

Figure 1.6. Promoter chromatin states on completely activated genes (Class I).

(A) Expression profiling on example Class I genes. S, SSCs. M, MASCs. E, ESCs.
(B) Chromatin states (PRIMs) of all Class I gene promoters.
(C) Comparison of chromatin states (PRIMs) between ESCs and MASCs for all class I gene promoters. X-axis, PRIMs from ESCs. Y-axis, PRIMs from MASCs. r, correlation coefficient for all class I gene promoter PRIMs.
(D) Comparison of chromatin states (PRIMs) between SSCs and MASCs for all class I gene promoters (grey dots). X-axis, PRIMs from SSCs. Y-axis, PRIMs from MASCs. Black dots, peak detection identified SSC bivalent genes. Red dots, Class I example genes displayed in Figure 1.6A. Dashed line, cut-off between activated (A) and repressed (R) chromatin states. r, correlation coefficient for all class I promoter PRIMs. Based on cut-off value, three groups of promoters were selected: $MASC^{Stable\ I}$, stably repressed in both SSCs and MASCs; $MASC^{Active}$, repressed in SSCs but activated in MASCs; $MASC^{M}$, not modified in SSCs but acquire either K4me3 or K27me3 modification in MASCs.
(E) Promoter modification at selected genes. Top, expression class and promoter type of each gene. Green, K4me3 modification. Red, K27me3 modification. K4me3 track range, 0 – 1. K27me3 track range, 0 – 0.5.

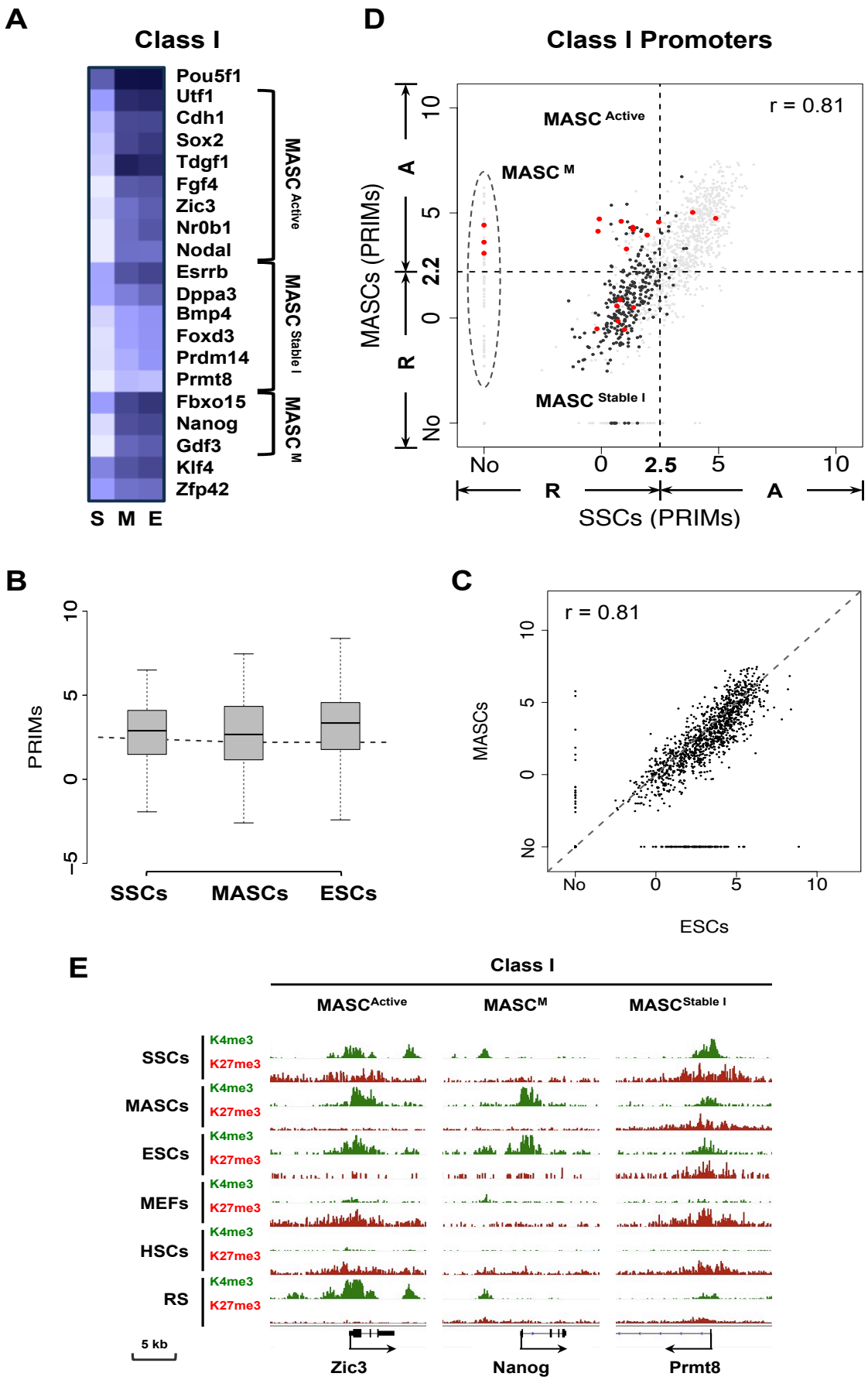
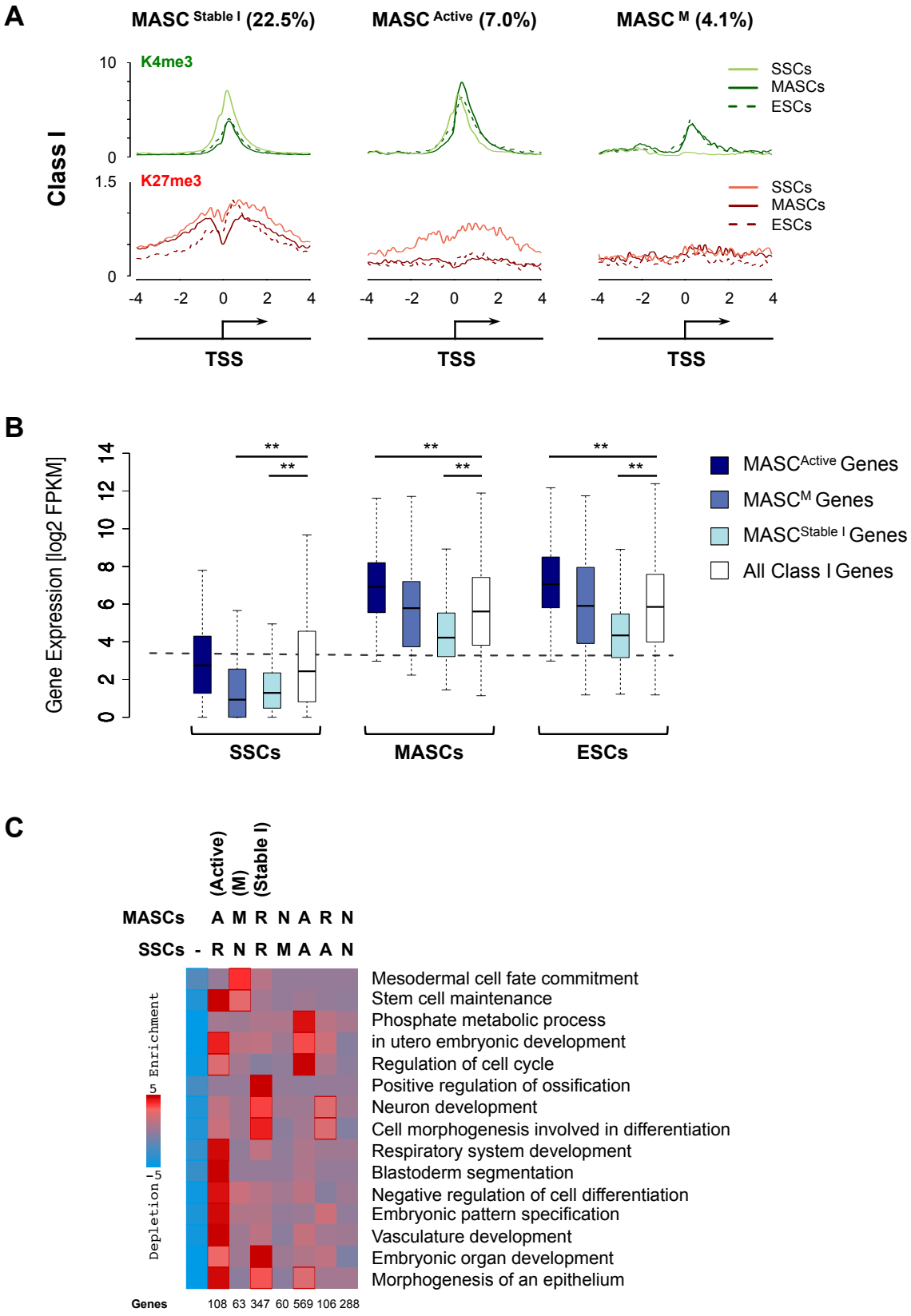

A
Class I
Pou5f1
Utf1
Cdh1
Sox2
Tdgf1
Fgf4
Zic3
Nr0b1
Nodal
Esrrb
Dppa3
Bmp4
Foxd3
Prdm14
Prmt8
Fbxo15
Nanog
Gdf3
Klf4
Zfp42
MASC Active
MASC Stable I
MASC M
S M E
D
Class I Promoters
r = 0.81
MASC Active
MASC M
MASC Stable I
MASCs (PRIMs)
SSCs (PRIMs)
No
R
A
2.2
B
PRIMs
SSCs
MASCs
ESCs
C
r = 0.81
MASCs
ESCs
E
Class I
MASCActive
MASCM
MASCStable I
SSCs
MASCs
ESCs
MEFs
HSCs
RS
K4me3
K27me3
5 kb
Zic3
Nanog
Prmt8

Figure 1.7. Promoter chromatin states, corresponding transcription activities, and biological functions on completely activated genes (Class I).

(A) Histone modification enrichment profiling for selected class I gene promoters (grouped as in Figure 1.6D). Percentage of genes in each class is listed in brackets. Solid green, K4me3 in SSCs. Solid dark green, K4me3 in MASCs. Dashed dark green, K4me3 in ESCs. Solid red, K27me3 in SSCs. Solid dark red, K27me3 in MASCs. Dashed dark red, K27me3 in ESCs. TSS, transcription start site. Arrow, direction of transcription. X-axis, distance to TSS. Y-axis, average read density.

(B) Transcription activities in selected subsets of Class I genes. Genes are grouped by promoter PRIMs as in Figure 1.6D. **, p-value < 0.01.

(C) GO enrichment using iPAGE. Genes are grouped by promoter PRIMs as in Figure 1.6D. A, active chromatin state. R, repressive chromatin state. M, detectable histone modification (K4me3 or K27me3). N, undetectable histone modification (K4me3 or K27me3). Bottom, number of genes in each group.


A
MASC Stable I (22.5%)
MASC Active (7.0%)
MASC M (4.1%)
Class I
K4me3
K27me3
SSCs
MASCs
ESCs
TSS
B
Gene Expression [log2 FPKM]
MASCActive Genes
MASCM Genes
MASCStable I Genes
All Class I Genes
SSCs
MASCs
ESCs
C
MASCs A M R N A R N
SSCs - R N R M A A N
(Active) (M) (Stable I)
Enrichment
Depletion
Genes 108 63 347 60 569 106 288
Mesodermal cell fate commitment
Stem cell maintenance
Phosphate metabolic process
in utero embryonic development
Regulation of cell cycle
Positive regulation of ossification
Neuron development
Cell morphogenesis involved in differentiation
Respiratory system development
Blastoderm segmentation
Negative regulation of cell differentiation
Embryonic pattern specification
Vasculature development
Embryonic organ development
Morphogenesis of an epithelium

Therefore, dynamic chromatin state changes were restricted to only a few ESC signature genes (Class I, $MASC^{Active}$ and $MASC^{M}$), indicating that promoter chromatin state-associated transcriptional activation is both selective and gene specific. However, genes functioning in embryonic differentiation to somatic lineages maintained their bivalent promoter modifications in both SSCs and MASCs, consistent with recent observations of freshly isolated mouse spermatogonia (Hammoud et al., 2014). Expression of such genes during SSC reprogramming could be regulated by mechanisms that do not affect promoter histone modifications, e.g., transcription factor binding at cell-type-specific enhancers.

4. Germ cell-specific genes (Class II) are repressed in MASCs with a general increase in promoter K27me3 modification

In contrast to Class I, Class II included 1,458 genes that were highly expressed in SSCs but down-regulated in MASCs and ESCs (Figure 1.5A, 1.8A). Correspondingly, most Class II promoters shifted from an active towards a more repressive chromatin state after reprogramming (Figure 1.8B). In particular, there were 40.9% promoters modified with concomitant *de novo* K27me3 and decreased K4me3 marks in MASCs (596 promoters, $MASC^{Repress}$), encompassing nearly three-fourths of Class II promoters that were epigenetically active in SSCs (Figure 1.8D, 1.9A, Table 1.6). Significant K27me3 increases were also observed at 14.8% promoters that preserved both repressed chromatin states and relatively low expression in both SSCs and MASCs (216 promoters, $MASC^{Stable\ II}$) (Figure 1.8D, 1.9A, B, Table 1.6). These transcriptionally poised genes ($MASC^{Repress}$ and $MASC^{Stable\ II}$) typically regulate embryonic differentiation toward germline and ectoderm lineages (e.g., Zbtb16/Plzf, Dmbx1) (Figure 1.8E, 1.9C). Furthermore, many spermatogenesis regulators exhibited

substantial reductions in the K4me3 mark, yielding unmodified promoter chromatin in MASCs (e.g., Piwil4) ($MASC^{NM}$) (Figure 1.8, 1.9, Table 1.6). $MASC^{NM}$ genes were generally silenced beyond the average Class II genes in MASCs (Figure 1.9B). Notably, many genes with promoter bivalent modifications in SSCs ($MASC^{Active}$, $MASC^{Stable\ I}$, $MASC^{Stable\ II}$) lost either K4me3 or both marks in other somatic stem or differentiated cells, but K4me3 modifications were typically preserved in round spermatids, the precursor of mature spermatozoa (Figure 1.8E).

Thus, the majority of somatic genes with bivalent promoter modifications in SSCs remained epigenetically repressed in MASCs, despite transcriptional activation ($MASC^{Stable\ I}$) or repression ($MASC^{Stable\ II}$) after reprogramming. However, most ESC signature genes and germ cell-specific genes underwent dynamic changes in promoter K27me3 modification, while a few of them either acquired or erased K4me3 modifications in MASCs. These two histone modifications significantly affected the promoter chromatin states of a small subset of transcriptionally activated genes (Class I), but more prominently affected the repressed genes (Class II).

Figure 1.8. Promoter chromatin states on completely repressed genes (Class II).

(A) Expression profiling on example Class II genes. S, SSCs. M, MASCs. E, ESCs.
(B) Chromatin states (PRIMs) of all Class II gene promoters.
(C) Comparison of chromatin states (PRIMs) between ESCs and MASCs for all class II gene promoters. X-axis, PRIMs from ESCs. Y-axis, PRIMs from MASCs. r, correlation coefficient for all class II gene promoter PRIMs.
(D) Comparison of chromatin states (PRIMs) between SSCs and MASCs for all class II gene promoters (grey dots). X-axis, PRIMs from SSCs. Y-axis, PRIMs from MASCs. Black dots, peak detection identified SSC bivalent genes. Red dots, Class II example genes displayed in Figure 1.8A. Dashed line, cut-off between activated (A) and repressed (R) chromatin states. r, correlation coefficient for all class II gene promoter PRIMs. Based on cut-off value, three groups of promoters are selected: $MASC^{Stable\ II}$, stably repressed in both SSCs and MASCs; $MASC^{Repress}$, activated in SSCs but repressed in MASCs; $MASC^{NM}$, lose all detectable modifications in MASCs.
(E) Promoter modification at selected genes. Top, expression class and promoter type of each gene. Green, K4me3 modification. Red, K27me3 modification. K4me3 track range, 0 – 1. K27me3 track range, 0 – 0.5.

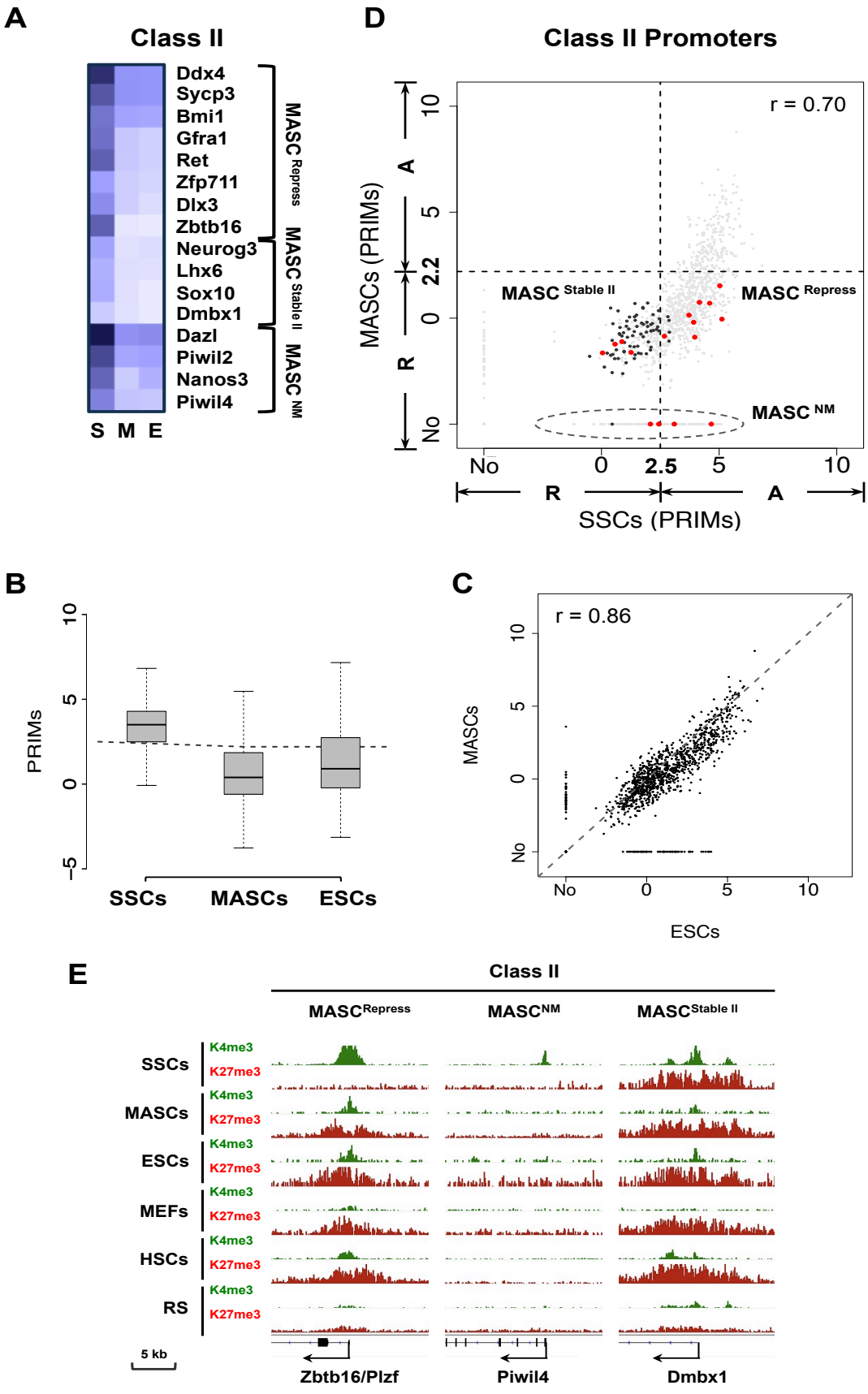
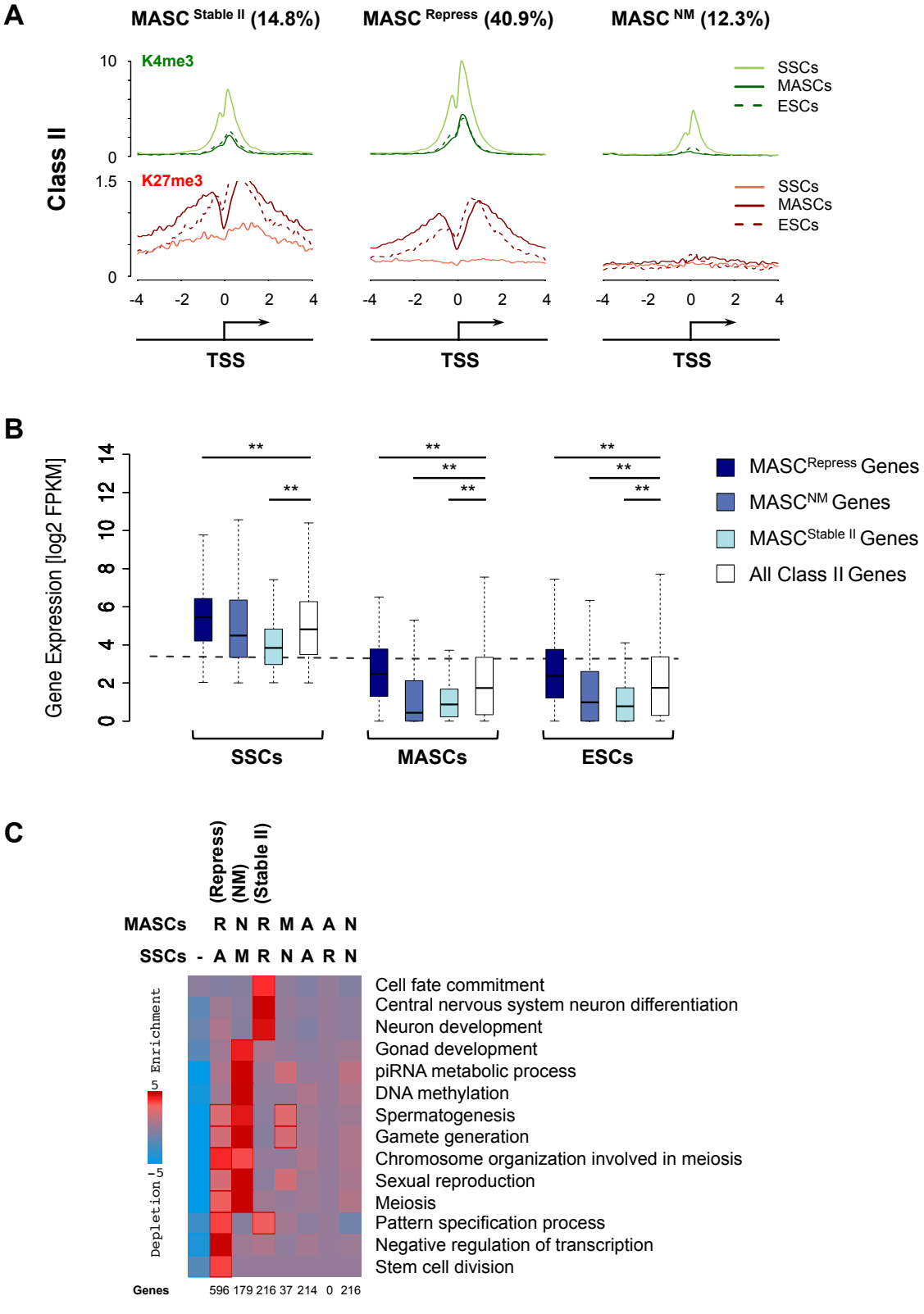

A
Class II
Ddx4
Sycp3
Bmi1
Gfra1
Ret
Zfp711
Dlx3
Zbtb16
Neurog3
Lhx6
Sox10
Dmbx1
Dazl
Piwil2
Nanos3
Piwil4
MASC Repress
MASC Stable II
MASC NM
S M E
D
Class II Promoters
r = 0.70
MASCs (PRIMs)
MASC Stable II
MASC Repress
MASC NM
No
R
A
SSCs (PRIMs)
B
PRIMs
SSCs
MASCs
ESCs
C
r = 0.86
MASCs
ESCs
E
Class II
MASCRepress
MASCNM
MASCStable II
SSCs
MASCs
ESCs
MEFs
HSCs
RS
K4me3
K27me3
5 kb
Zbtb16/Plzf
Piwil4
Dmbx1

Figure 1.9. Promoter chromatin states, corresponding transcription activities, and biological functions on completely repressed genes (Class II) in SSCs, MASCs, and ESCs.

(A) Histone modification enrichment profiling for selected class II gene promoters (grouped as in Figure 1.8D). Percentage of genes in each class is listed in brackets. Solid green, K4me3 in SSCs. Solid dark green, K4me3 in MASCs. Dashed dark green, K4me3 in ESCs. Solid red, K27me3 in SSCs. Solid dark red, K27me3 in MASCs. Dashed dark red, K27me3 in ESCs. TSS, transcription start site. Arrow, direction of transcription. X-axis, distance to TSS. Y-axis, average read density.

(B) Transcription activities in selected subsets of Class II genes. Genes are grouped by promoter PRIMs as in Figure 1.8D. **, p-value < 0.01.

(C) GO enrichment using iPAGE. Genes are grouped by promoter PRIMs as in Figure 1.8D. A, active chromatin state. R, repressive chromatin state. M, detectable histone modification (K4me3 or K27me3). N, undetectable histone modification (K4me3 or K27me3). Bottom, number of genes in each group.


A
MASC Stable II (14.8%)
MASC Repress (40.9%)
MASC NM (12.3%)
Class II
K4me3
K27me3
SSCs
MASCs
ESCs
TSS
B
Gene Expression [log2 FPKM]
MASCRepress Genes
MASCNM Genes
MASCStable II Genes
All Class II Genes
SSCs
MASCs
ESCs
C
(Repress)
(NM)
(Stable II)
MASCs R N R M A A N
SSCs - A M R N A R N
Enrichment
Depletion
Genes
596 179 216 37 214 0 216
Cell fate commitment
Central nervous system neuron differentiation
Neuron development
Gonad development
piRNA metabolic process
DNA methylation
Spermatogenesis
Gamete generation
Chromosome organization involved in meiosis
Sexual reproduction
Meiosis
Pattern specification process
Negative regulation of transcription
Stem cell division

5. Incompletely reprogrammed genes (Class III, IV) and MASC-specific genes (Class V, VI) achieve ESC-like promoter chromatin states

In addition to completely reprogrammed genes, we also identified genes that stayed silent (Class III) (Figure 1.5A, 1.10A, B) or were consistently expressed (Class IV) (Figure 1.5A, 1.10E, F) in both SSCs and MASCs in contrast to ESCs (Table 1.6). Each class constituted only ~3% genes in this study. Surprisingly, these incompletely reprogrammed genes shared similar promoter chromatin states between MASCs and ESCs (Figure 1.10C, D, G, H). In particular, the K27me3 modification was significantly enriched at Class IV promoters in both MASCs and ESCs but not in SSCs, suggesting that these promoters switched from an active to a poised chromatin state prior to completion of transcriptional repression (Figure 1.10G, H).

An ESC-like promoter chromatin state was also observed at most MASC-specific overexpressed (Class V) (Figure 1.11A – D) or silenced (Class VI) genes (Figure 1.11E – H). Notably, many mesoderm-specific genes (e.g., Snai1, Class IV; Sox7, Class V) and epithelial-specific genes (e.g., Gata6, Class V) acquired not only high expression but also bivalent promoter modifications in MASCs (Figure 1.10E, I, 1.11A, I), indicating that transcription of such genes could be easily adjusted by a change in chromatin modification.

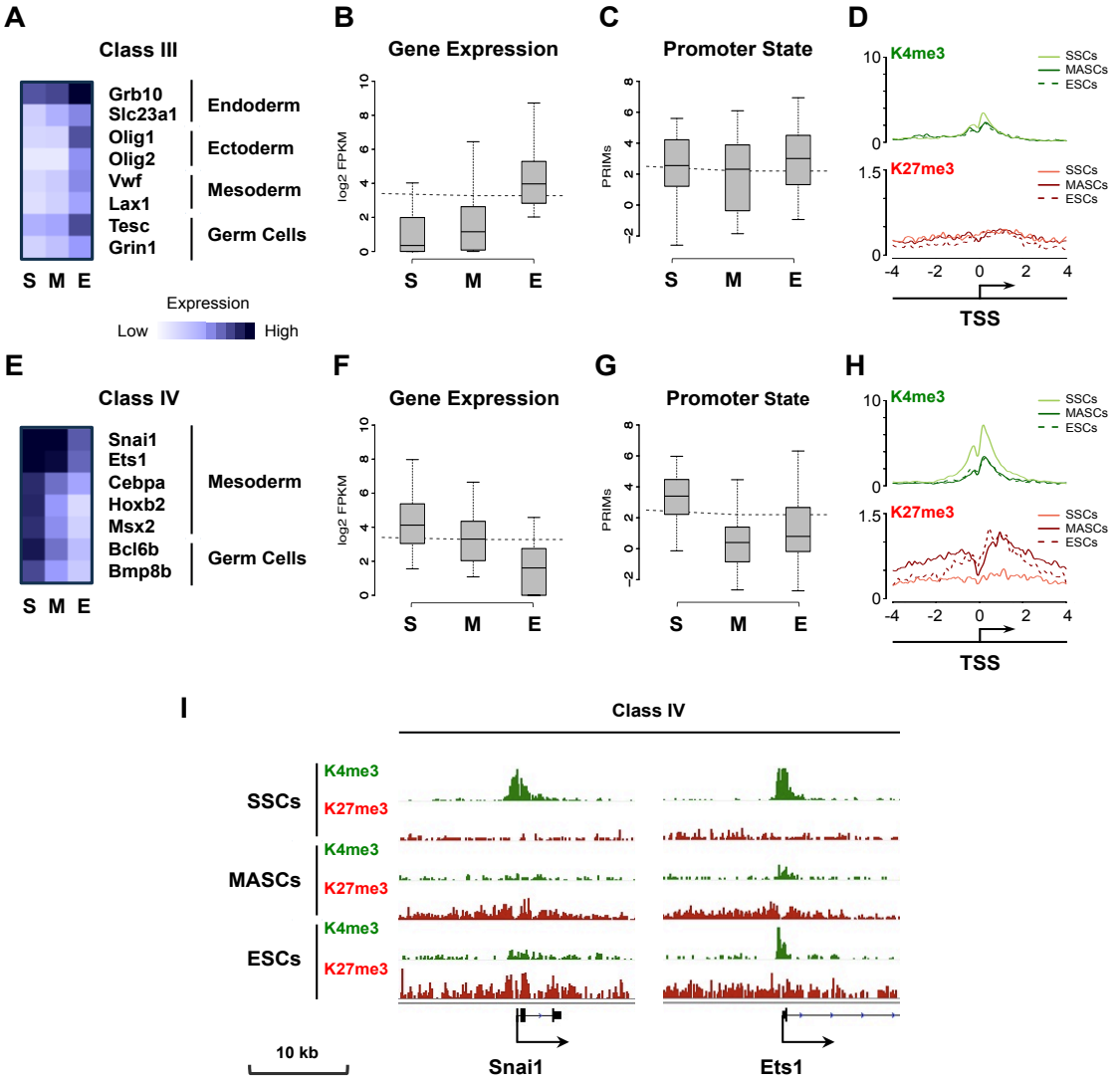


Figure 1.10. Incompletely reprogrammed genes (Class III, IV) achieve ESC-like promoter chromatin states.

Expression profiling on example (A) Class III or (E) Class IV genes among SSCs (S), MASCs (M), and ESCs (E). Associated differentiation lineages are shown to the right.

Expression profiling of all (B) Class III or (F) Class IV genes. Dashed line: average expression of all genes in each cell type.

Promoter PRIMs profiling of all (C) Class III or (G) Class IV genes. Dashed line: average promoter PRIMs of all genes in each cell type.

Histone modification enrichment profiling for all (D) Class III or (H) Class IV gene promoters.

(I) Promoter modification at selected genes. Top, expression class and promoter type of each gene. Green, K4me3 modification. Red, K27me3 modification. K4me3 track range, 0 – 1. K27me3 track range, 0 – 0.5.

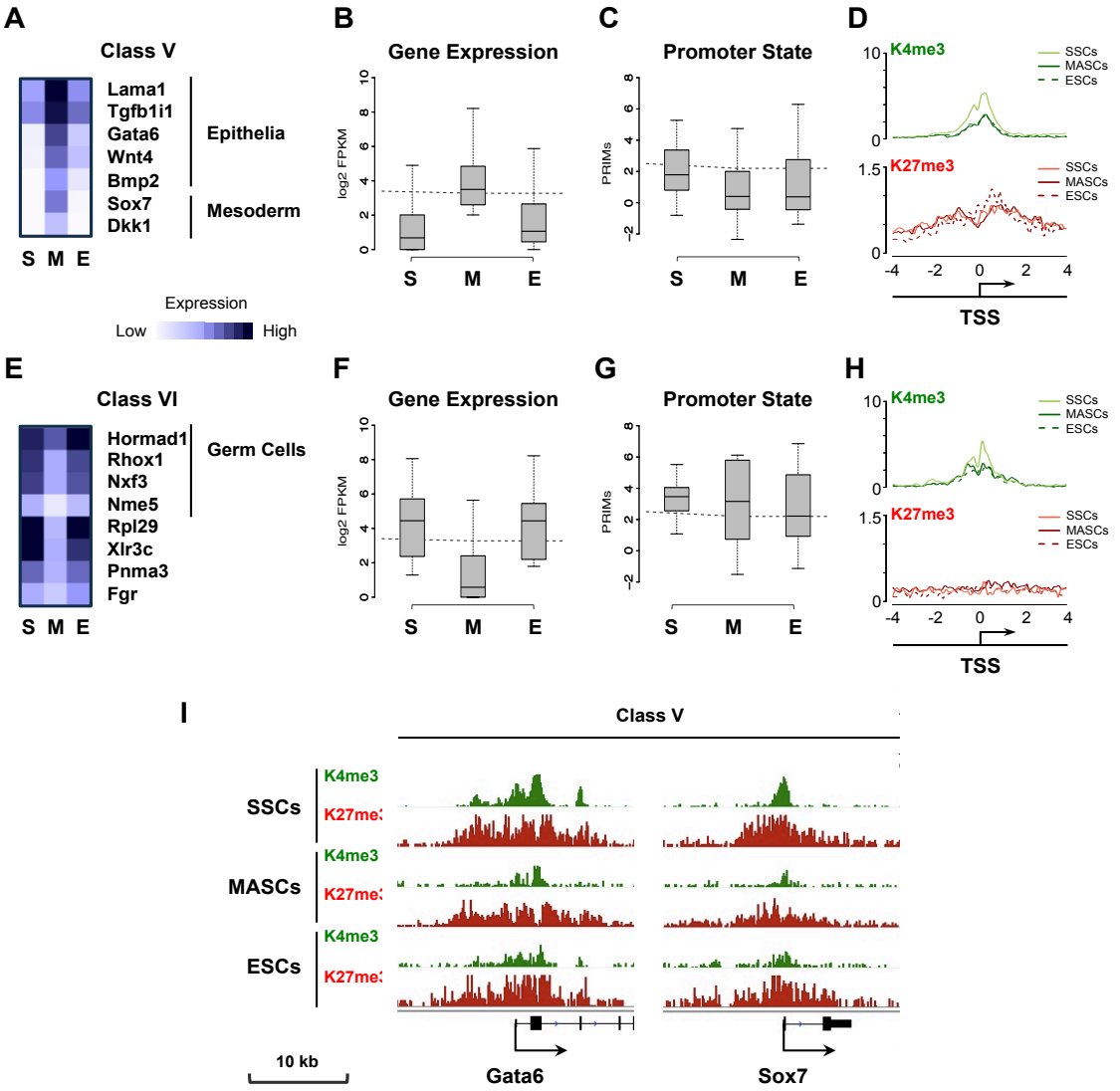


Figure 1.11. MASC-specific genes (Class V, VI) achieve ESC-like promoter chromatin states.

Expression profiling on example (A) Class V or (E) Class VI genes among SSCs (S), MASCs (M), and ESCs (E). Associated differentiation lineages are shown to the right.

Expression profiling of all (B) Class V or (F) Class VI genes. Dashed line: average expression of all genes in each cell type.

Promoter PRIMs profiling of all (C) Class V or (G) Class VI genes. Dashed line, average promoter PRIMs of all genes in each cell type.

Histone modification enrichment profiling for all (D) Class V or (H) Class VI gene promoters.

(I) Promoter modification at selected genes. Top, expression class and promoter type of each gene. Green, K4me3 modification. Red, K27me3 modification. K4me3 track range, 0 – 1. K27me3 track range, 0 – 0.5.

6. MASCs are depleted of germ cell-specific enhancer activity but partially enriched for ESC enhancers

To investigate the mechanisms of incomplete reprogramming (Class III, IV) and MASC-specific gene expression (Class V, VI), we extended our analysis of epigenetic regulation beyond promoters by using H3K27ac (K27ac) ChIP-seq, which detected active enhancers defined as regions with K27ac peaks but without known promoters or K4me3 peaks (K27ac+/K4me3-).

A total of 9,738 enhancers were found to be active in at least one of the three cell types (SSCs, MASCs, ESCs) and were associated with the 3,316 differentially expressed genes (Table 1.7). As predicted, MASCs, having lost lineage commitment, shared more active enhancers with pluripotent ESCs (1,582, ME enhancers) than with unipotent SSCs (194, SM enhancers) (Figure 1.12A, B). Of note, one-third of enhancers activated in both MASCs and ESCs were also active in SSCs (803, SME enhancers) (Figure 1.12A). DNA motif search with the HOMER computer program revealed that many ESC signature transcription factors (e.g., Sox2, Nr5a2) could bind at these common active enhancers, indicating that an ESC-like transcriptional program might be partially active in SSCs (Figure 1.13, Table 1.8). Conversely, the 3,107 ESC-specific enhancers were targeted by many late embryogenesis regulators (e.g., Usf1/2, Foxh1), suggesting that the ESC differentiation apparatus could be responsible for activation of such enhancers (Figure 1.13). Study of ChIP-seq data sets in ESCs confirmed that ME and SME enhancers near the ESCs active genes (e.g., Zic3, Nanog in Class I; Lax1 in Class III) were highly occupied by many ESC signature transcription factors, particularly those centered at pluripotency circuitry (Pou5f1/Oct4, Sox2, Nanog) (Figure 1.14, 1.15, Table 1.7). However, not many ESC- or SSC-specific enhancers were targets of core pluripotency transcription factors

(Figure 1.14, 1.15). Only Esrrb and Tfcp2l1 were found at ESC-specific enhancers (e.g., Gata6, Class V), consistent with their roles in establishing gold standard, ESC-defined pluripotency (Figure 1.14, 1.15).

There were little unique motif enrichment at the 1,336 MASC-specific enhancers, except for Atf3, Prdm14, Smad3, and NFkB signaling factors, which could play roles in MASC formation (Figure 1.13, Table 1.8). Considering their relative dominance near Class V genes, MASC-specific enhancers could promote overexpression of those genes (Figure 1.12C). Furthermore, we also identified a group of enhancers active in SSCs that were silent in MASCs but active in ESCs (552; referred to as SE enhancers); many were associated with genes suggestive of incomplete reprogramming in MASCs (e.g., Snai1, Class IV) (Figure 1.15). Motifs enriched at both SSC-specific enhancers (e.g., ETS, FOX families) and ESC-specific enhancers (e.g., Zfx) were identified in these SE enhancers (Figure 1.13, Table 1.8); some SE enhancers were also occupied by core pluripotency transcription factors in ESCs (Figure 1.15). Such a pattern could indicate that a switch from germ cell-type transcriptional regulation to pluripotency circuitry occurs concomitant with SSC reprogramming and that incomplete establishment of ESC-specific enhancer activity could delay reprogramming at a small percent of genes.

In summary, MASCs exhibited erasure of germline-specific enhancers and partially established enhancers that resemble those of ESCs. This process was prospectively enabled by binding of ESCs signature transcription factors. This switch of enhancer activity closely correlated with changes in expression of nearby genes and could support SSC reprogramming along with changes in promoter chromatin state. ESC-specific enhancers that remained silent in MASCs would not affect cell differentiation to different lineages.

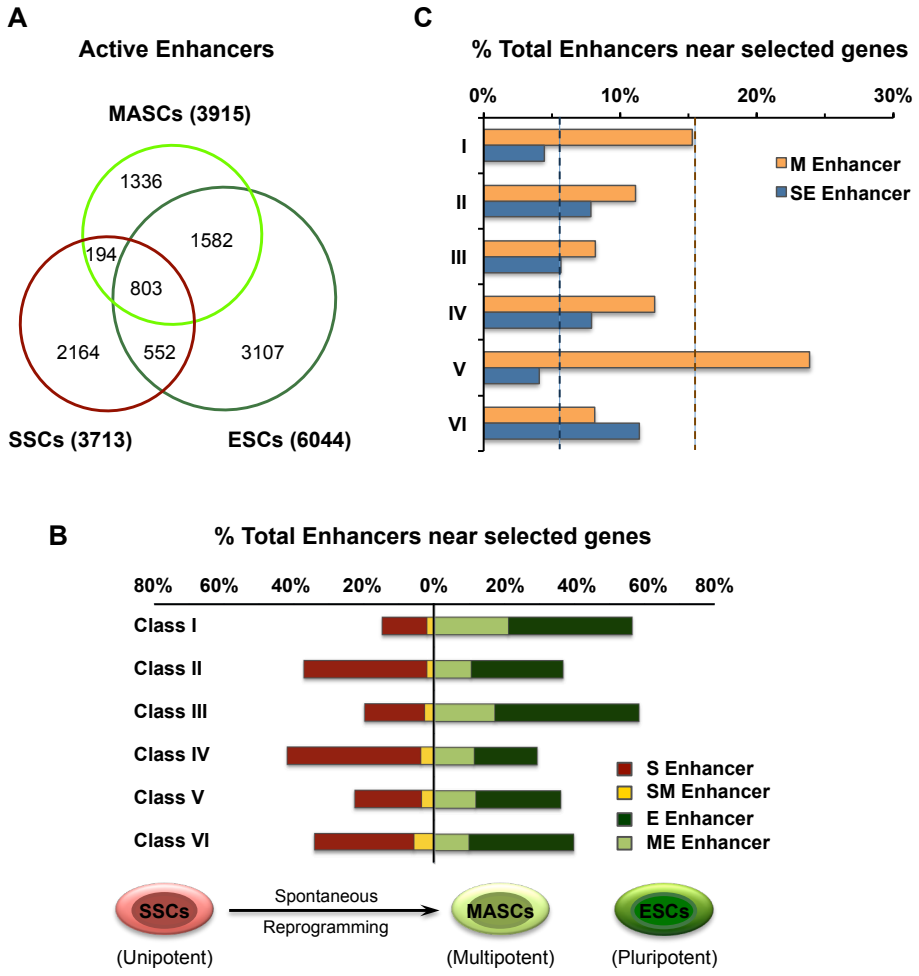


Figure 1.12. Enhancer activity at different classes of genes in SSCs, MASCs, and ESCs.

(A) Overlap of active enhancers in SSCs, MASCs, and ESCs.

(B) Percentage of enhancers near Class I ~ VI genes. S enhancer, enhancers active only in SSCs. SM enhancer, enhancers active in SSCs and MASCs. E enhancer, enhancers active only in ESCs. ME enhancer, enhancers active in MASCs and ESCs.

(C) Percentage of enhancers near Class I ~ VI genes. M enhancer, enhancers active only in MASCs. SE enhancer, enhancers active in SSCs and ESCs. Brown dashed line, percentage of M enhancer near all genes. Blue dashed line, percentage of SE enhancer near all genes.

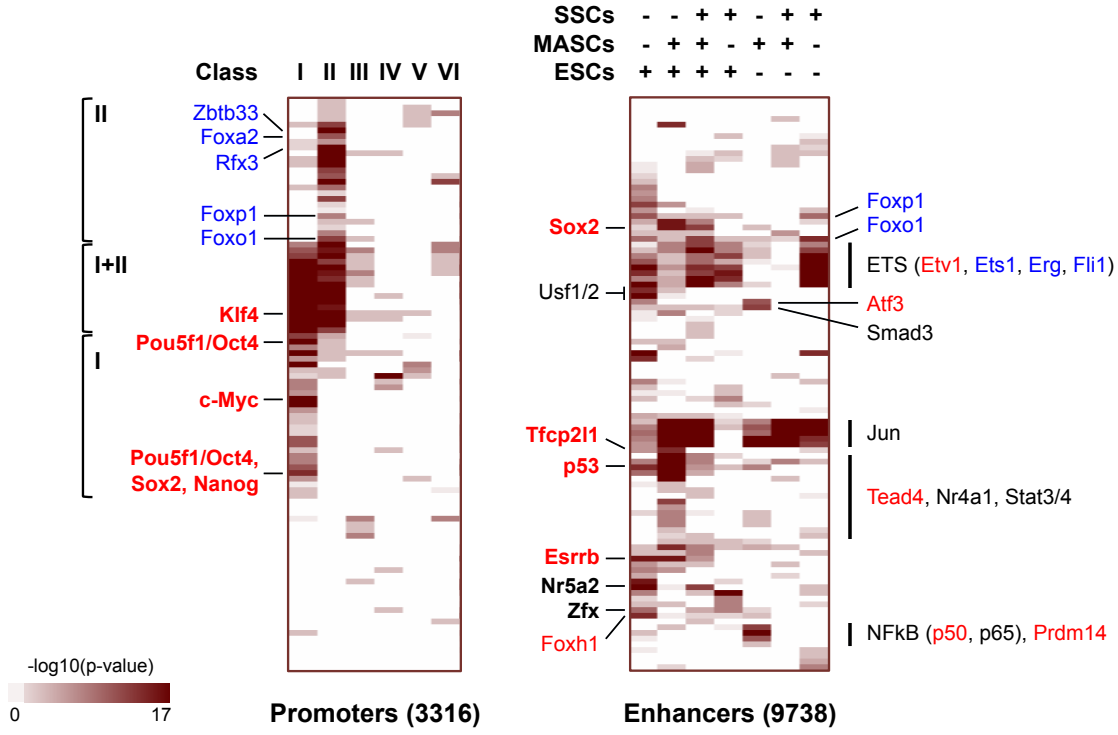


Figure 1.13. Enrichment of transcription factor binding motifs for class I ~ VI gene promoters (left) and enhancers active (+) or silence (-) in different cell types (right). Each row represents a motif, and the corresponding transcription factors are labeled on the side. Red, genes up-regulated (Class I, III) in ESCs. Blue, genes down-regulated (Class II, IV) in ESCs.

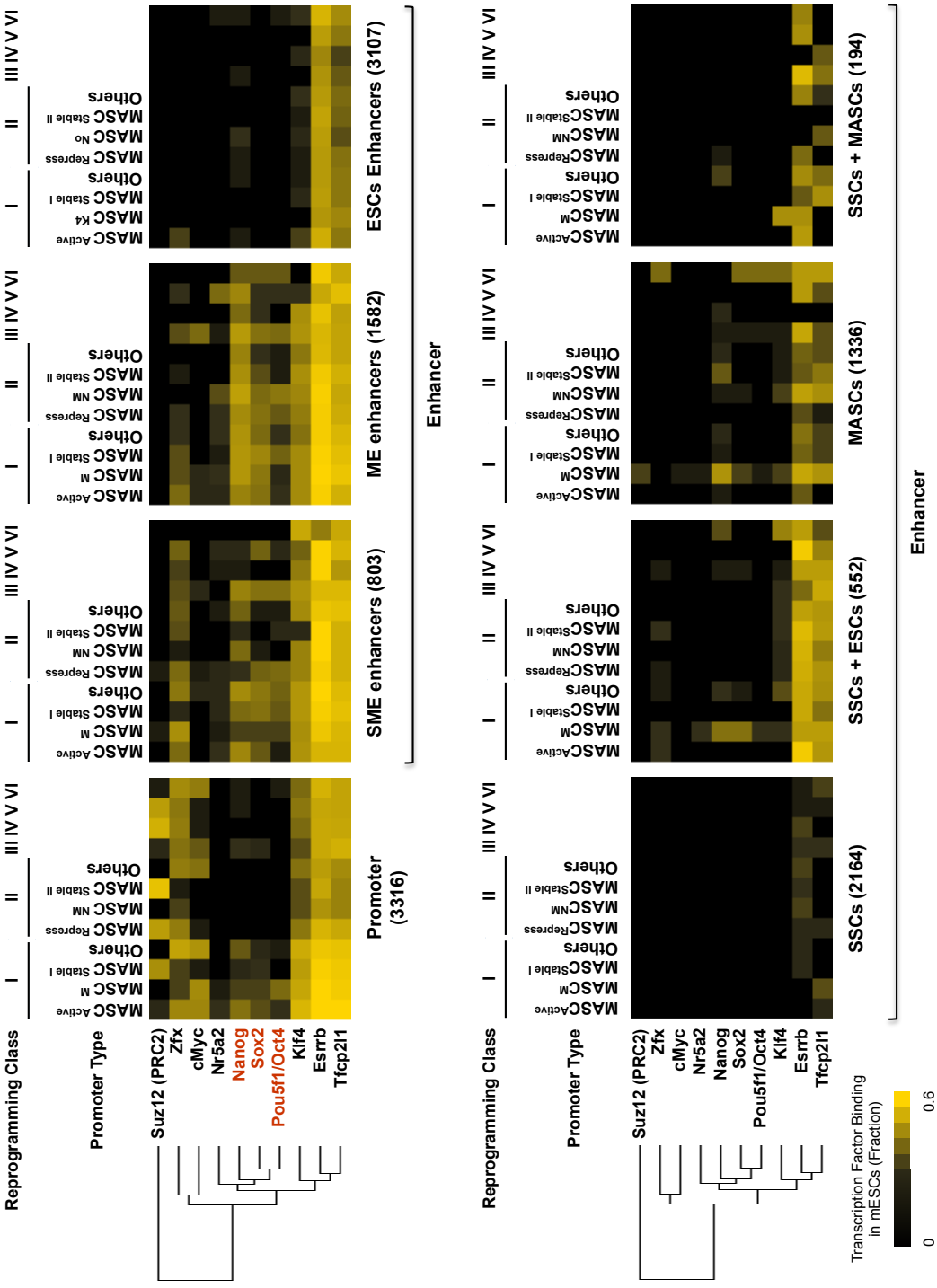


Figure 1.14. Enrichment of transcription factors at promoters and associated enhancers in ESCs. Promoters were grouped by expression class as in Figure 1.5A and chromatin state as in Figure 1.6D, 1.8D. Dark orange, core pluripotency transcription factors.

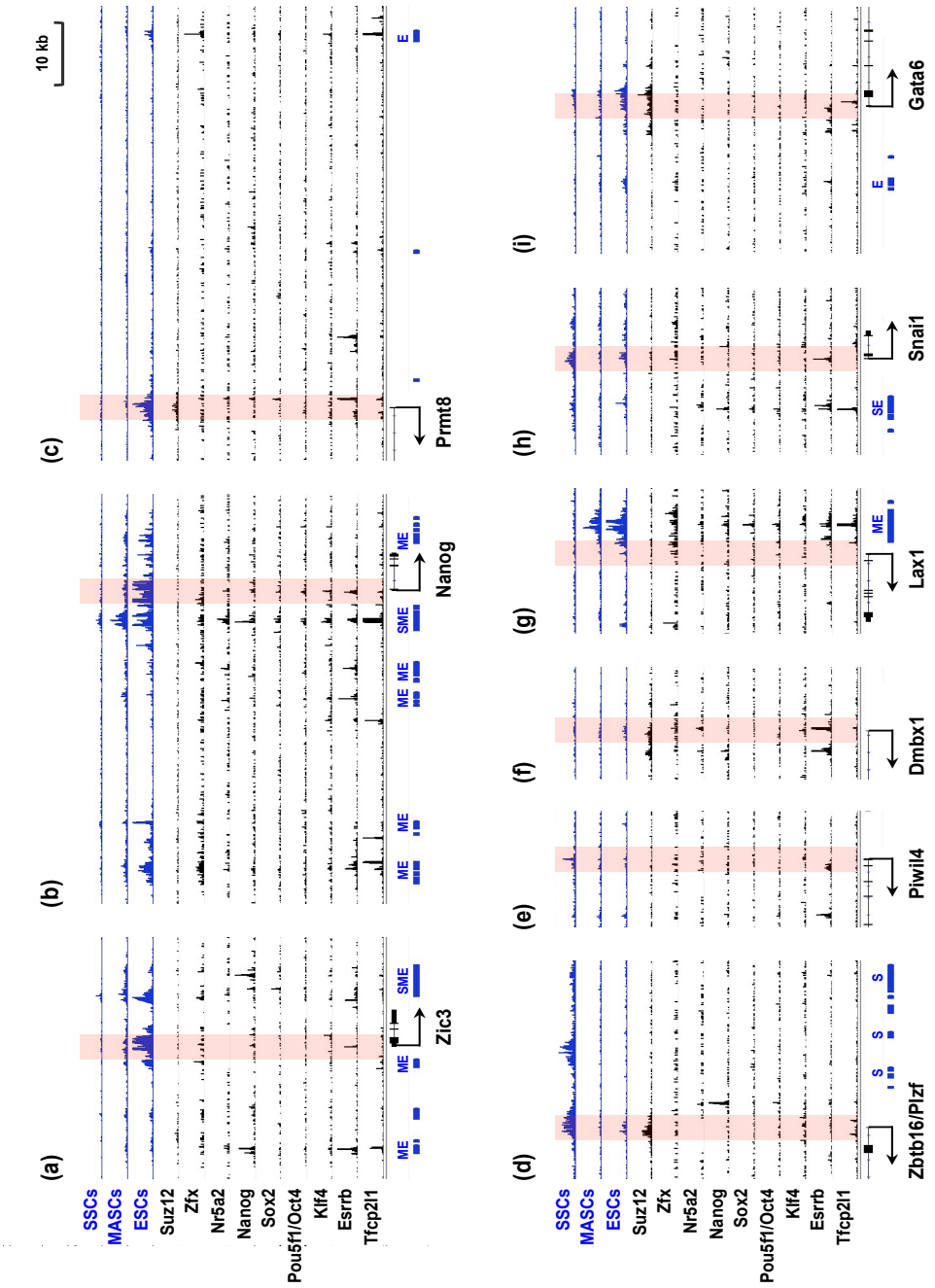


Figure 1.15. K27ac modification (blue) and transcription factor enrichment in ESCs (black) at selected genes. a, Zic3 (Class I, MASC^{Active}). b, Nanog (Class I, MASC^{M}). c, Prmt8 (Class I, MASC$^{Stable\ I}$). d, Zbtb16 (Class II, MASC^{Repress}). e, Piwil4 (Class II, MASC^{NM}). f, Dmbx1 (Class II, MASC$^{Stable\ II}$). g, Lax1 (Class III). h, Snai1 (Class IV). i, Gata6 (Class V). Red region, promoters. Blue bar at bottom, enhancer region. E, enhancers activated only in ESCs. ME, enhancers activated in both MASCs and ESCs. SME, enhancers activated in SSCs, MASCs and ESCs. SE, enhancers activated in both SSCs and ESCs. S, enhancers activated only in SSCs. All track range, 0 – 1.

7. Expression of somatic genes with promoter bivalency is potentially balanced by PRC2 (Suz12) and cell-type-specific transcription factors

To elucidate the molecular mechanisms involved in transcriptional regulation and changes in promoter chromatin state, we also applied DNA motif search with HOMER to the transcriptionally-defined promoter classes. Notably, transcriptionally activated promoters (Class I) were specifically enriched for motifs of many Class I stem cell pluripotency regulators (e.g., Pou5f1/Oct4, Sox2, Nanog, c-Myc), whereas repressed promoters (Class II) were preferential targets of ESC differentiation regulators that lost expression in MASCs (e.g., FOX family, Rfx3, Class II) (Figure 1.13, Table 1.8). However, promoters of class III ~ VI were not significantly enriched for such motifs (Figure 1.13). This analysis indicated that completely reprogrammed genes could be auto-regulated by their own classes of transcripts. This notion was supported by comparison with published ChIP-seq data sets, showing that core pluripotency transcription factors generally bind to completely activated Class I promoters in ESCs (Figure 1.14, 1.15) (Chen et al., 2008; Heng et al., 2010). Therefore, we suggest that SSC reprogramming requires recruitment of distinct transcription factors that control gene activation and repression at promoter regions and that this regulation is likely critical to complete reprogramming.

Analysis of ChIP-seq data also revealed that the PRC2 subunit Suz12 was enriched at promoters with K27me3 modification in ESCs. Most of these promoters lacked binding motifs of core pluripotency transcription factors, except for some somatic genes that maintained promoter bivalency but were up-regulated in MASCs (e.g., Prmt8, Class I, $MASC^{Stable\ I}$) (Figure 1.14, 1.15). Binding of core pluripotency transcripts at epigenetically poised promoters could improve the expression of these

somatic genes without altering their underlying promoter chromatin states. By contrast, very few core pluripotency transcription factors were found at bivalent promoters that were down-regulated in MASCs and ESCs (e.g., Dmbx1, Class II, $MASC^{Stable\ II}$) (Figure 1.14, 1.15). Instead, such bivalent promoters were targets of multiple regulators of embryonic differentiation, many of which show decreased expression in MASCs (e.g., ETS, FOX families) (Figure 1.13). Taken together, these data suggest that transcription of down-regulated bivalent promoters ($MASC^{Stable\ II}$) are controlled by such differentiation factors that are expressed in SSCs and become silent during SSC reprogramming, thereby down-regulating target promoters in MASCs. In MASCs. Polycomb-mediated repression could dominate the promoters of such lineage-specific genes in MASCs and reinforce the developmental flexibility.

In summary, these findings suggest that bivalently modified somatic gene promoters could be regulated by both PRC2(Suz12) and target-specific transcription factors and that their cooperation drives active transcription or silencing of these genes, respectively, in a cell-type-specific manner.

8. SSCs retain promoter bivalency comparable with germ cells *in vivo*.

Recently, genome-wide studies of the germline epigenome revealed that the K4me3+K27me3 bivalent histone modification remains faithfully stable from the development of embryonic progenitors through to postnatal germ cells (Lesch et al., 2013; Ng et al., 2013; Sachs et al., 2013). Besides germ cell-specific genes, many developmental regulators functioning in somatic lineages are poised with both K4me3 and K27me3 histone modifications at promoters in PGCs (Lesch et al., 2013; Ng et al., 2013), adult germline stem cells (AGSCs) (Hammoud et al., 2014), pachytene

spermatocytes (Lesch et al., 2013; Hammoud et al., 2014), round spermatids (Erkek et al., 2013; Lesch et al., 2013; Hammoud et al., 2014), and mature spermatozoa (Hammoud et al., 2014). This epigenetic feature is therefore suggested to be essential for germ cell identity and function (Lesch and Page, 2014). To understand whether long-term *in vitro* cultured SSCs preserve similarly poised chromatin as do germ cells *in vivo*, we compared K4me3 and K27me3 ChIP-seq results from cultured SSCs with those of published data.

We focused on 2,347 SSC bivalent genes, which were selected by peak detection as genes with significant K4me3 and K27me3 marks at promoters in at least two SSC cell lines (Table 1.5). Cluster analysis using PRIM values for these promoters showed that, cultured SSCs not only shared similar K4me3- and K27me3-defined chromatin states with cultured MASCs and ESCs but also with PGCs directly isolated from embryonic gonads (E12.5 – E14.5) (Lesch et al., 2013). However, multiple types of progenitors or differentiating germ cells isolated from adult testes were relatively distant from SSCs, similar as somatic stem or differentiated cells applied in this study (Figure 1.16A). We next investigated dynamic changes in histone modifications at each promoter within the germline lineage. With *k*-means clustering function in seqMINER (Ye et al., 2011), SSC bivalent genes were grouped according to similarity of promoter K4me3 and K27me3 profiles in PGCs (E12.5, E13.5, E14.5), AGSCs, pachytene spermatocytes, round spermatids, and mature spermatozoa (Figure 1.16B, Table 1.5). Intriguingly, there were 358 genes (cluster I) that maintained relatively high K4me3 and K27me3 signals from PGCs to spermatozoa (Figure 1.16B, C). The increase in K27me3 modification was initiated as early as E13.5 in PGCs, the time point of sex determination in the embryonic gonad, accelerated in postnatal testes, and persisted at significant levels in spermatozoa (Figure 1.16C). Furthermore, this class of genes was significantly enriched for developmental regulators that direct the

differentiation of both germ and somatic cells, including Sox2, Cdx2, and Gata6, among others. (Figure 1.16D, E). The K4me3 and K27me3 patterns on these developmental genes remained stable in both embryonic and postnatal germ cells, but some of them resolved to K4me3 or K27me3 monovalent chromatin in somatic cells (MEFs) (Figure 1.16E). Importantly, these germline-poised developmental genes shared similar K4me3 and K27me3 modification patterns with cultured SSCs, providing support for the concept that germ cell identity in SSC cell lines is maintained long-term *in vitro*.

Figure 1.16. SSCs maintain consistent promoter bivalency with germ cells *in vivo*.

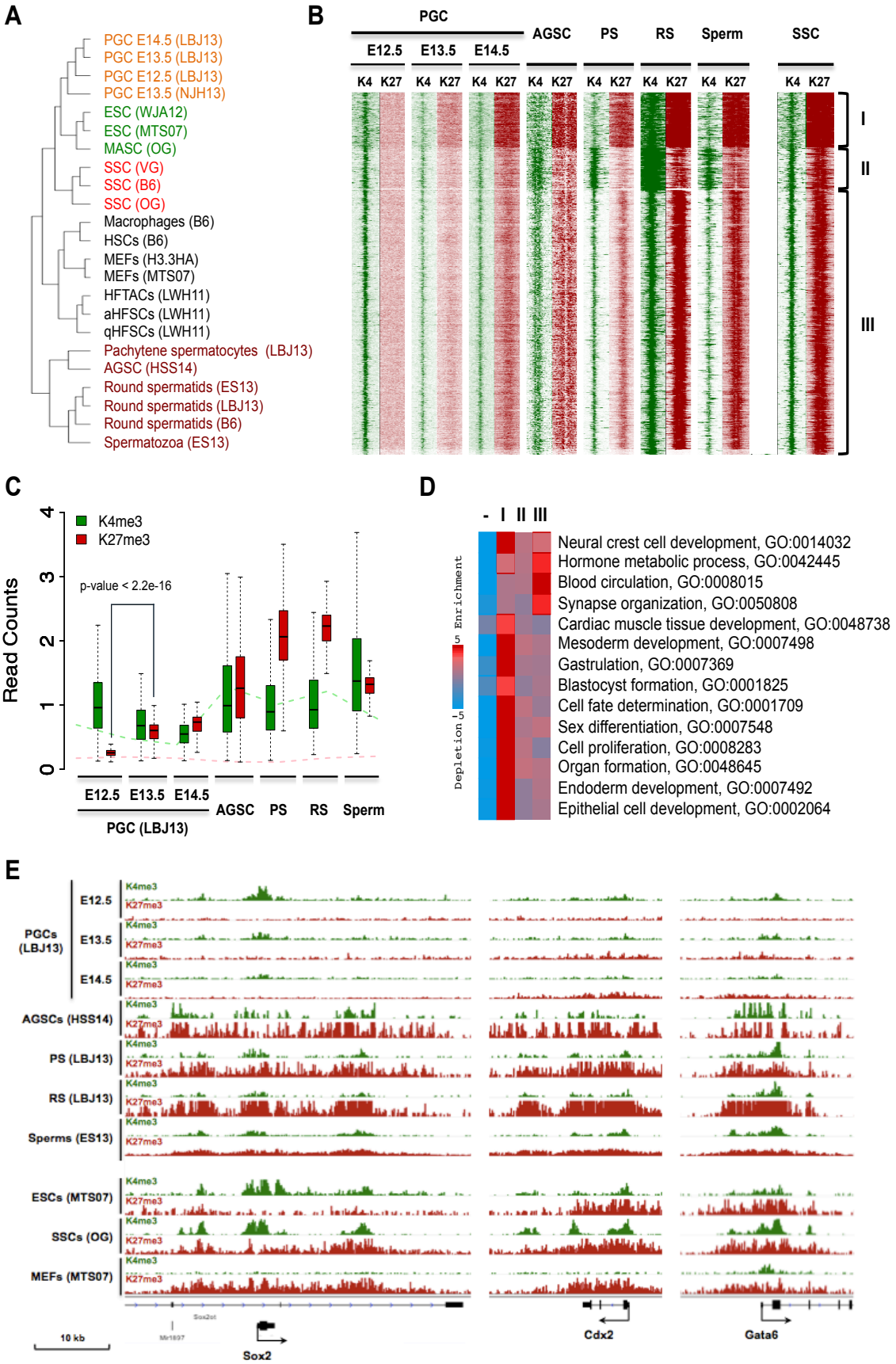
(A) Dendrogram based on PRIMs of all promoters with K4me3+K27me3 bivalent histone modification in SSCs. Cells with the same or similar function were labled with the same color. Green, pluripotent ESCs (Mikkelsen et al., 2007; Wamstad et al., 2012) and MASCs. Orange, PGCs collected at different developmental stages (Lien et al., 2011; Ng et al., 2013). Red, cultured SSCs (Hammoud et al., 2014). Dark red, pre- or post-meiotic germ cells isolated from adult testes (Lien et al., 2011; Erkek et al., 2013). Black, somatic cells, including MEFs (Mikkelsen et al., 2007), quiescent (q-)/activated (a-) hair follicle stem cells (HFSCs), hair follicle transient-amplifying matrix cells (HFTACs) (Lien et al., 2011), Hematopoietic stem cells (HSCs), and Macrophages. ChIP-seq data source or strain information is listed in brackets. Mouse strain includes B6, OG, VG, H3.3HA.

(B) *k*-means clustering of genes by similarity of K4me3 and K27me3 profiles at SSC bivalent gene promoters. Green, K4me3. Red, K27me3. Cluster I, 358 genes. Cluster II, 276 genes. Cluster III, 1713 genes. PS, pachytene spermatocyte. RS, round spermatid.

(C) Histone modification profiling at promoters of all cluster I genes as grouped in Figure 1.16B. Y-axis, average read count within promoter region. Green box, K4me3 modification. Red box, K27me3 modification. Dashed green line, average read count of K4me3 modification at all promoters in each cell type. Dashed red line, average read count of K27me3 modification at all promoters in each cell type.

(D) GO enrichment using iPAGE. Genes were grouped by promoter K4me3 and K27me3 profiles as in Figure 1.16B.

(E) Histone modification at selected genes. Green, K4me3 modification. Red, K27me3 modification. K4me3 track range, 0 – 1. K27me3 track range, 0 – 0.5.


A
PGC E14.5 (LBJ13)
PGC E13.5 (LBJ13)
PGC E12.5 (LBJ13)
PGC E13.5 (NJH13)
ESC (WJA12)
ESC (MTS07)
MASC (OG)
SSC (VG)
SSC (B6)
SSC (OG)
Macrophages (B6)
HSCs (B6)
MEFs (H3.3HA)
MEFs (MTS07)
HFTACs (LWH11)
aHFSCs (LWH11)
qHFSCs (LWH11)
Pachytene spermatocytes (LBJ13)
AGSC (HSS14)
Round spermatids (ES13)
Round spermatids (LBJ13)
Round spermatids (B6)
Spermatozoa (ES13)
B
PGC
E12.5
E13.5
E14.5
AGSC
PS
RS
Sperm
SSC
K4
K27
I
II
III
C
K4me3
K27me3
p-value < 2.2e-16
Read Counts
0
1
2
3
4
E12.5
E13.5
E14.5
AGSC
PS
RS
Sperm
PGC (LBJ13)
D
-
I
II
III
Enrichment
5
Depletion
5
Neural crest cell development, GO:0014032
Hormone metabolic process, GO:0042445
Blood circulation, GO:0008015
Synapse organization, GO:0050808
Cardiac muscle tissue development, GO:0048738
Mesoderm development, GO:0007498
Gastrulation, GO:0007369
Blastocyst formation, GO:0001825
Cell fate determination, GO:0001709
Sex differentiation, GO:0007548
Cell proliferation, GO:0008283
Organ formation, GO:0048645
Endoderm development, GO:0007492
Epithelial cell development, GO:0002064
E
PGCs (LBJ13)
E12.5
E13.5
E14.5
AGSCs (HSS14)
PS (LBJ13)
RS (LBJ13)
Sperms (ES13)
ESCs (MTS07)
SSCs (OG)
MEFs (MTS07)
K4me3
K27me3
10 kb
Sox2
Cdx2
Gata6

9. Histone variant H3.3 is preserved at core pluripotency genes in SSCs.

In addition to histone modificaitons, several histone variants are also suggested to play important roles in transcription regulation and cell development. Recently, genome-wide study of H3.3 localization in mammalian post-meiotic germ cells has shown that H3.3 is preserved at many developmental loci in haploid round spermatids and mature spermatozoa (Erkek et al., 2013). This finding raises an intriguing hypothesis that H3.3 could carry and transfer epigenetic information between generations, and its developmental gene-specific localizaiton in paternal genome potentially indexes the activation of these genes upon fertilizaiton.

Activation of key pluripotency genes is a critical step in cell reprogramming. To elucidate the potential role of H3.3 in SSC reprogramming, we investigated genome-wide H3.3 enrichment in cultured SSCs by ChIP-seq. SSCs were derived from adult mouse that carries a HA epitope tag at the C-terminal of endogenous H3f3b gene (H3.3-HA). After *in vitro* expansion under standard SSC culture condition, cells were collected and subjected to ChIP with antibody against HA epitope and then submitted for sequencing. Besides, H3.3 ChIP-seq results were also collected from ESCs and MEFs from the same mouse strain. For comparisons, we reprocessed data from published datasets for haploid round spermatids and mature spermatozoa (Erkek et al., 2013).

To evaluate the sequence signal of H3.3 at every gnomic locus in each cell type, the whole mouse genome was segmented into non-overlapping 1kb windows, and H3.3 sequence read count per window was calculated by SNPseeqer developed in Dr. Elemento's lab. Interestingly, correlation clustering with genome-wide H3.3 enrichment indicates that H3.3 distribution in SSCs is relatively similar to ESCs than

to MEFs. Conversely, post-meiotic germ cells including round spermatids and spermatozoa are more distant from the other deploid cells, presumabably due to the global histone-to-protamine exchange in round spermatids (Figure 1.17A).

By examining H3.3 localization in SSCs, we noticed that most genes that are expressed in ESCs but silent in SSCs are not enriched with H3.3. However, H3.3 is significantly enriched at both gene bodies and regulatory regions (promoters, enhancers) of the three well-known core pluripotency genes Pou5f1/Oct4, Nanog, and Sox2 in SSCs (Figure 1.17B). Although RNA-seq showed that these three ESC-specific transcription factors are barely expressed in SSCs, the H3.3 distribution patterns are surprisingly similar between ESCs and SSCs. Conversely, H3.3 does not occupy the promoters and gene bodies of Sox2 and Nanog in differentiated MEFs, but remains modestly detectable at these pluripotency genes in haploid round spermatids and mature spermatozoa (Figure 1.17B).

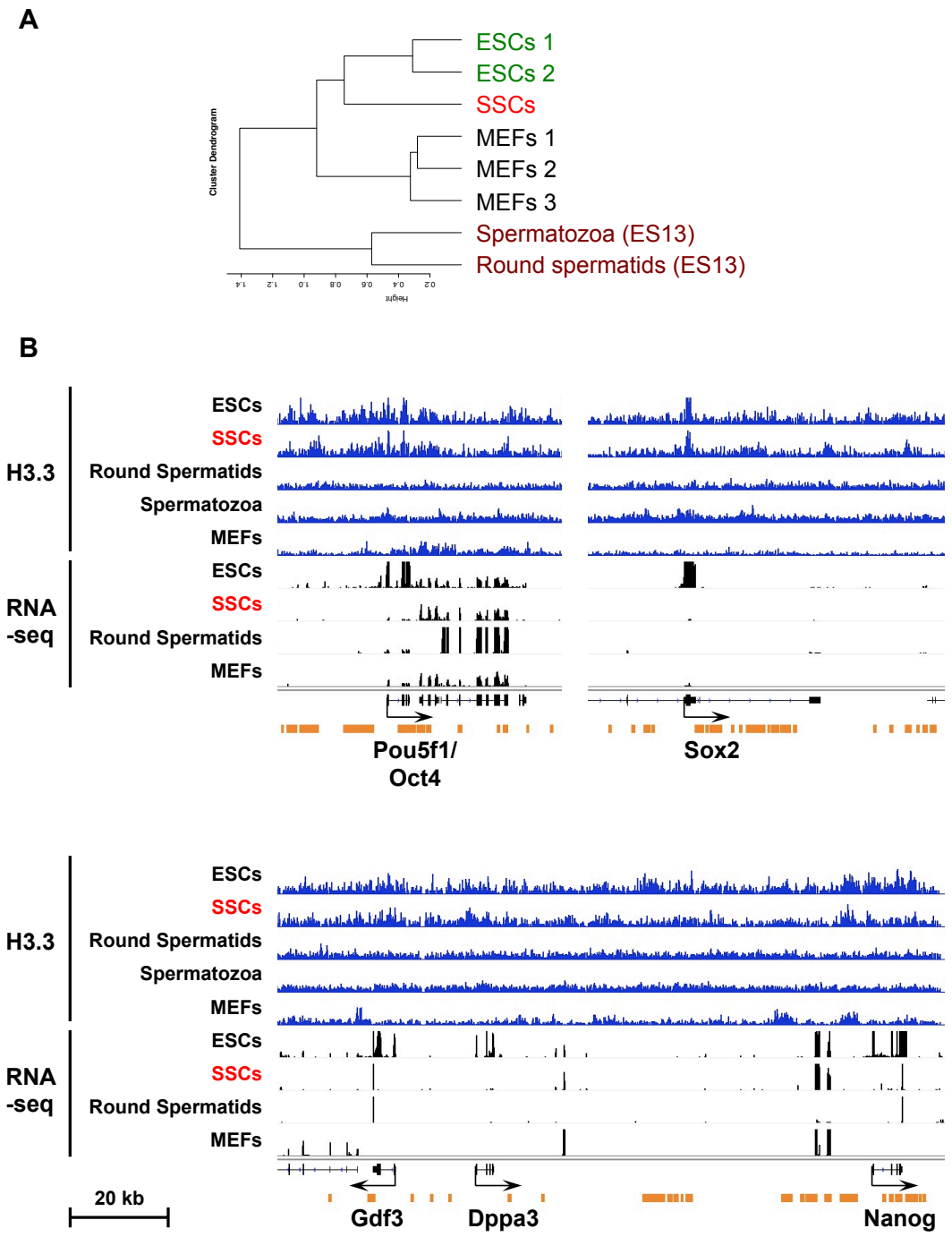


Figure 1.17. H3.3 is preserved at core pluripotency genes in SSCs.

(A) Dendrogram based on genome-wide H3.3 enrichment profiles.

(B) H3.3 enrichment at selected genes. Blue, H3.3. Black, RNA-seq result. Orange, ESC-specific enhancer region. Same data range is applied to all cell types for H3.3 tracks (0 – 0.3) and RNA-seq tracks (0 – 2.0), except for H3.3 tracks for round spermatids and spermatozoa (0 – 0.15).

IV. Discussion

Here, we evaluate the genomic localization of several transcription-associated histone modifications in mouse SSCs and their reprogrammed counterparts (MASCs). In depth analyses of the SSC epigenome at transcription regulatory loci, coupling with corresponding gene expression profiling, reveal that while the somatic gene promoters maintain stable K4me3+K27me3 bivalent histone modification, promoters and enhancers of ESC and germ cell signature genes undergo significant epigenetic changes after SSC reprogramming into a multipotent state. This gene-specific epigenetic conversion not only reflects the transcriptome difference between SSCs and MASCs, but also provides several insights about the unique biological characteristics in SSCs. And the SSC epigenome could play a critical role in both cell reprogramming *in vitro* and differentiation *in vivo* (Figure 1.18).

1. SSC conversion is an unique reprogramming event.

Somatic cells (i.e. MEFs) reprogramming to induced pluripotent stem cells (iPS) can be initiated by overexpression of several key pluripotency-associated transcription factors. A stepwise process has been proposed to take place during the switch from lineage-defined cells to pluripotent ESC-like cells (Samavarchi-Tehrani et al., 2010; Buganim et al., 2012). This reprogramming process could include the transcription events transfer from early stochasticity to more deterministic and hierarchic later phase in reprogramming cells, as it has been suggested by the marker gene expression changes at different reprogramming stages (Buganim et al., 2012). According to this model, overexpression of key pluripotency-associated transcription factors stochastically induces gene expression and initiates reprogramming, and the random activation of endogenous pluripotency genes progressively establishes ESC-like

hierarchic transcriptional regulation, which eventually stablizes the pluripotent stem cell identity at the end of this cell fate transition.

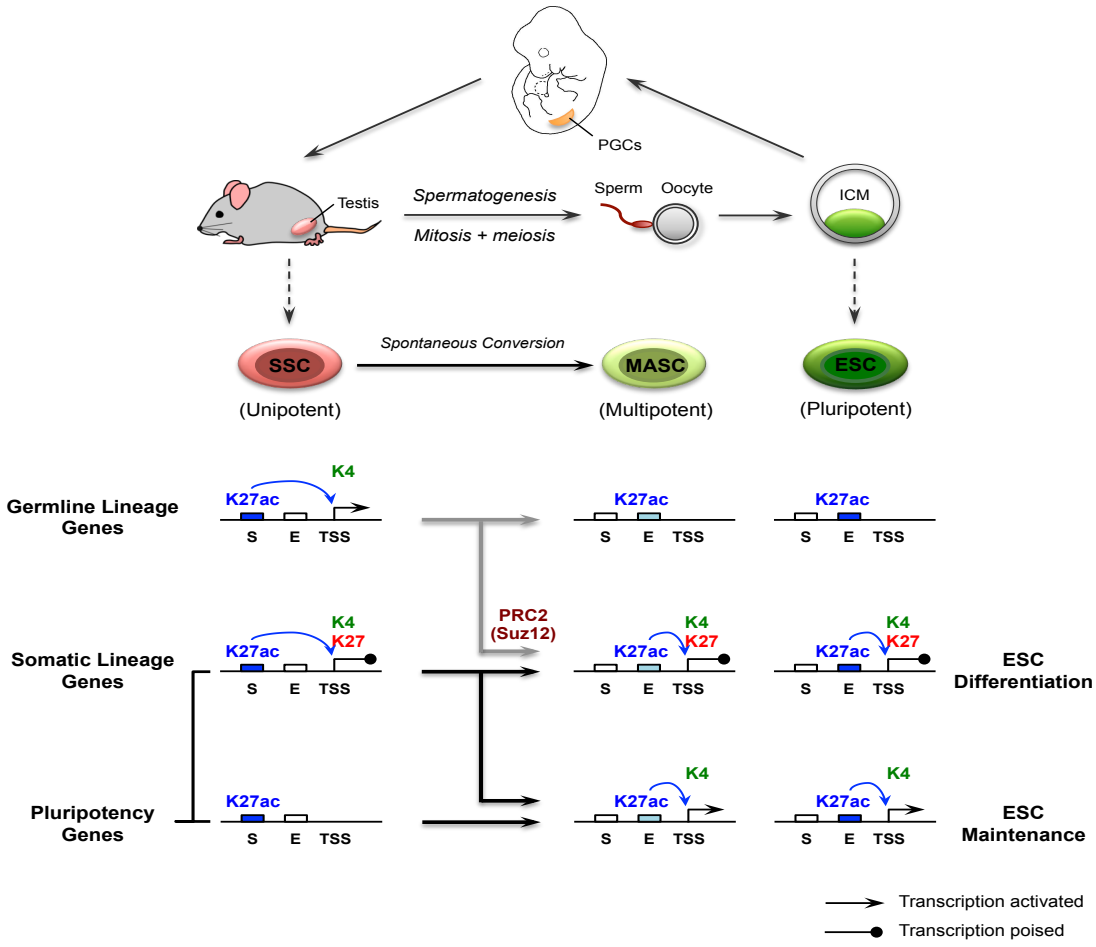


Figure 1.18. Model of epigenome changes during SSC conversion to MASC.

SSC conversion involves change of epigenome at both cell-type-specific promoters and enhancers. Together with gene reactivation and silencing in reprogrammed MASCs, K4me3 modification (K4) is enriched at the promoters of pluripotency-associated genes but depleted from promoters of germline-specific genes. Conversely, K27me3 modification (K27) is erased from several pluripotency-associated gene promoters. However, K27me3 modification 'poises' both germline- and somatic-specific gene promoters together with K4me3 modification. Many of these (K4+K27) bivalent modified promoters are targets of PRC2 (Suz12) in pluripotent ESCs. At enhancer regions, MASCs are erased of most germ cell signature enhancer activity (S) marked by H3K27ac modification, but partially activate ESC-specific enhancers (E). In summary, MASCs acquire ESC-like epigenome at promoters and a subset of ESC-specific enhancers. Dark blue box, completely activated enhancer. Light blue box, incompletely activated enhancer. White box, silent enhancer.

Conversely, considering the nature of SSC conversion, which happens spontaneously in standard SSC culture condition without any exogenous gene induction, the multipotent MASCs could be derived from a reprogramming process distinct from iPS formation. We proposed that intrinsic genetic and epigenetic features rather than enforced expression of ectopic transcription factor(s) are responsible for SSC conversion. In our model, SSCs are poised to tri-lineage differentiation by ESC-like genetic or epigenetic factors in their own genome. This intrinsic developmental flexibility provides SSCs the chance of retrieving multipotent developmental characteristics during long-term proliferation *in vitro*. However, It is still unknown what induces SSC reprogramming process. Either a slight change of culture environment or progressive accumulation of unknown reprogramming initiation factors could be responsible for the cell fate transition from defined germline lineage to all three somatic germ layers.

2. Transcriptome change involved in SSC conversion.

Transcriptome analysis reveals that MASCs not only activate somatic cell- and ESC-expressed genes but also repress germ cell differentiation regulators. During early embryogenesis, many of these activated genes are expressed in the inner cell mass (ICM), the embryonic origin of ESCs, and are considered marks of stem cell pluripotency. This transcriptiome change implies that SSCs reach a ESC-like developmental state through recovery of a broad range of embryonic developmental programs and erasure of adult germ cell characteristics. Both processes could be critical to the achievement of somatic lineage differentiation potential in MASCs. Our model suggests that intrinsic biological features are responsible for SSC conversion. Therefore, the transition from germline to embryonic transcriptome is presumably well

orchestrated under the SSC genome and/or epigenome, rather than stochastically induced as being suggested during somatic cell reprogramming to iPS.

My intensive study on epigenome shows that somatic gene promoters are poised for transcription in SSCs by both K4me3 and K27me3 histone modifications. Despite of significant expression changes after SSC conversion, these genes maintain bivalent promoter modifications in MASCs ($MASC^{Stable\ I}$ and $MASC^{Stable\ II}$). This poised promoter chromatin state potentially provides flexible transcriptional switch on somatic genes, and activation of these embryonic developmental programs could be one of the early events that initiates SSC conversion.

However, my study also shows that the promoters and enhancers of these epigenetically poised genes are potentially regulated by both PRC2(Suz12)-mediated K27me3 and by transcription factors, including the core pluripotency,regulators Pou5f1/Oct4, Sox2, and Nanog. Many of these transcription factors undergo significant expression increase (Class I) and promoter histone modification changes during SSC conversion. These results indicate that a few ESC-specific transcription factors could be on top of somatic genes in the reprogramming hierarchy. That is, epigenetic switches at the promoters of key transcription factors could be essential to their transcription activation and the initiation of SSC conversion, while somatic genes represent downstream targets of this hierarchical network.

On the other hand, silence of adult germ cell-specific genes could also initiate SSC conversion. Unlike somatic genes, promoters of these germline-expressed genes (Class II) mainly accumulate with K27me3 or lose K4me3 in MASCs, indicating that epigenetic regulation potentially provide direct impact on the silence of these germ cell-specific genes after SSC conversion. Given that germline-specific gene products potentially dominate over and repress somatic genes in SSCs, silence of these genes

could release the repression on stably poised genes ($MASC^{Stable\ I}$ and $MASC^{Stable\ II}$) in SSCs and initiate their expression in MASCs.

Because SSC conversion is a rare and long-term event without established induction protocol, it is difficult to identify the initiation factors of this process with current techniques. We were unable to complete time-series study on reprogramming SSCs, especially at the initiation stage of the reprogramming process, making it hard to study the first step of SSC transcriptome transition. Although histone modification change potentially provides significant impact on expression increase of key pluripotency genes and decrease of spermatogenesis genes, it is hard to exclude the possibility that epigenetically poised somatic genes initiate SSC conversion through direct response to the reprogramming-associated signaling. Identification of the regulators that promote the transcriptome transition will be essential to understand the mechanism of SSC conversion and to improve the reprogramming efficiency.

We also noticed that several mesenchymal and epithelial genes are specifically expressed in MASCs (Class V). Their promoters are poised with bivalent modifications in both SSCs and MASCs. It raises the possibility that the completion of SSC conversion involves a mesenchymal-epithelial transition (MET), which has been proposed to be a critical step in initiating iPS formation (Samavarchi-Tehrani et al., 2010). However, it is unknown whether MET gene expression is a stable feature of MASCs, or it is just a consequence of MASCs differentiation to mesenchymal and epithelial lineages. Further examination of MASCs transcriptome with improved culture condition (i.e., 2i medium) will be necessary to address this quesiton.

Taking together, time-series study on the entire SSC conversion process, together with in-depth analyses of the transcriptome and corresponding epigenome, will be necessary to understand the mechanism of this unique cell reprogramming event.

3. Stable pluripotency-associated epigenetic signatures underlying germline development.

Study of epigenome shows that SSCs preserve ESC-like epigenetic features at both promoters and enhancers. In contrast to the differentiated cells that have been analyzed, promoters of embryonic differentiation-associated genes maintain K4me3+K27me3 bivalent modifications in both SSCs and ESCs. Moreover, several ESC signature genes are also poised with bivalent modifications at promoters and histone variant H3.3 in gene body in SSCs. Besides, motif study of active enhancers shared among SSCs, MASCs, and ESCs suggests that core pluripotency circuitry is partially active in SSCs. This unique epigenetic milieu could facilitate the transcriptional switch in the presence of an internal or external stimulus and endow SSCs with unusual developmental flexibility with respect to all lineages in addition to the germline.

Recently, it has been reported that a set of developmental genes are poised with K4me3+K27me3 bivalent modification in mamalian germline, from PGCs, the embryonic progenitors of SSCs, to mature spermatozoa (Lesch et al., 2013; Ng et al., 2013; Sachs et al., 2013). Our findings in SSCs further confirm the developmental plasticity in the adult germ cells. Besides, our data shows that long-term *in vitro* cultured SSCs preserve chromatin modifications at selected promoters as do progenitor cells *in vivo*. The amazing chromatin consistency on developmental gene promoters suggests that germline epigenome could ensure stable transfer of epigenetic 'memory' to the next generation, despite of global epigenomic changes in embryonic gonad and during meiosis. Further study on these genes with germline-stable and -specific epigenetic marks will shed light on our understanding of the initiation of embryogenesis.

4. How to improve SSC conversion?

a. Improving SSC epigenome by small molecules in culture

After SSC conversion, MASCs are depleted of germ cell-specific epigenetic signatures at most *cis*-regulatory regions. Instead, MASCs acquire nearly completely ESC-like promoter chromatin modifications but partially activate pluripotent ESC-specific enhancers. Because MASCs have the capability of differentiating to all germ layers during teratoma formation, promoter histone modifications, together with enhancers shared between MASCs and ESCs, could be sufficient to support ESC-like tri-lineage differentiation. Those enhancers that remain silent in MASCs are presumably not essential to the core pluripotency circuitry, but could be induced during long-term culture or with improved growth condition.

However, SSC reprogramming process takes over months and is about 100-fold less efficient than generating iPS from MEFs. Besides, MASCs are known to be inefficient in both chimera contribution and germline transmission, and completely lack tetraploid complementation ability (Kanatsu-Shinohara et al., 2004; Seandel et al., 2007). As a result, it will still be important to investigate whether the establishment of complete embryonic-like enhancer activity can improve SSC conversion and MASC developmental potency to multiple lineages. One possible solution is to improve the K27ac-defined epigenome in culture by histone deacetylases (HDACs) inhibitors (i.e., VPA) (Huangfu et al., 2008). Development of novel, transcription factor-free, enforced reprogramming strategies will greatly benefit stem cell application in the clinic, and will also improve our understanding on the origin of totipotency during development.

b. Potential role of K4 and K27 methyltransferase/demethylase in SSC conversion

After SSC conversion, most pluripotency gene promoters switch to active chromatin states with erasure of K27me3 modifications ($MASC^{Active}$) or acquisition of the K4me3 modifications ($MASC^{M}$). Conversely, promoters of germ cell differentiation genes are mainly repressed with K27me3 modifications ($MASC^{Repress}$) or silenced with loss of K4me3 ($MASC^{NM}$). Besides, many incompletely reprogrammed promoters (Class III, IV) achieve ESC-like chromatin states, indicates that epigenetic conversion on these genes presumably takes place prior to actual transcriptional changes. It suggests that H3K4- and H3K27- specific methytransferases and demethylases may play active roles in SSC conversion, for example, by facilitating chromatin state changes at selected promoters. Further study of enzymes that mediate H3K4 and H3K27 covalent modification, particularly those specifically expressed in SSCs, could be necessary to improve reprogramming efficiency.

c. Potential role of H3.3 in SSC conversion

Genome-wide study of H3.3 enrichment in cultured SSCs reveals that H3.3 is highly preserved at the three core pluripotency-associated transcription factors, Pou5f1/Oct4, Nanog, and Sox2. This unique deposition can not be simiply explained by transcription activity, as RNA-seq shows that these three genes are barely transcribed in SSCs. Therefore, it raises an intriguing hypothesis that H3.3, as K4me3+K27me3 bivalent modification, could poise these genes for activation in SSCs. This leads to my in-depth study of H3.3 activity in cell reprogramming (Chapter two). To dissect the initiation stage of general cell reprogramming process, the transcription factor-induced iPS formation from MEFs was used as a model system.

CHAPTER TWO

HISTONE VARIANT H3.3 IS ENGAGED IN PLURIPOTENCY-ASSOCIATED GENE ACTIVATION DURING MOUSE EMBRYONIC FIBROBLAST REPROGRAMMING

I. Summary

To investigate the role of H3.3 during mammalian cell reprogramming, we evaluate the dynamic changes of transcriptome and corresponding H3.3-defined epigenome during MEFs reprogramming to transcription factor-induced pluripotent stem cells (iPS). Time-series analyses suggest that global H3.3 replacement initiates at the early stage of MEFs reprogramming (first ten days). Notably, a H3.3-signified ESC-like chromatin landscape is established prior to corresponding pluripotent gene activation. Furthermore, H3.3 replacement not only associates with K27me3 reduction at the promoters and enhancers of ESC signature genes, but also with re-establishment of K4me3+K27me3 bivalent modifications at embryonic differentiation gene promoters. Our study suggests that, upon reprogramming initiation, H3.3 is a novel epigenetic regulator that is involved in pluripotency gene activation through remodeling differentiated cell chromatin. This result, together with the notion that H3.3 is pre-deposited at key pluripotent genes in unipotent SSCs, shed light on our study of epigenetic regulation during cell reprogramming.

II. Introduction

Series of programming events orchestrate mammalian development from a fertilized zygote to an adult individual. While differentiated cells can also be reprogrammed to pluripotency by nuclear transfer, cell fusion, over-expression of a set of defined transcription factors, or spontaneous reversion (SSC culture only) (Takahashi and Yamanaka, 2006; Seandel et al., 2007; Wen et al., 2014b).

Histone H3 variant H3.3 has been implicated in chromatin remodeling and epigenetic regulation of gene expression during various mammalian cell development processes (Banaszynski et al., 2010). In the early mouse zygote, maternal H3.3 plays an important role in male pronucleus formation before the transcription activation in parental pronuclei (van der Heijden et al., 2005; Torres-Padilla et al., 2006). H3.3 deposition into the paternal genome is suggested to be critical for the establishment of pericentric heterochromatin, which ensures proper chromosome segregation during the first mitosis (Santenard et al., 2010). Accumulating evidences implicate that H3.3 also functions in germline development during adulthood (Banaszynski et al., 2010). Significant level of H3.3 has been identified in differentiating germ cells (Figure 10). Recent genome-wide studies further suggest that H3.3 is the dominant H3 histone in post-meiotic germ cells, including round spermatids and mature spermatozoa. The H3.3-nucleosomes, together with K4me3 and K27me3 modifications, enrich at a broad range of developmental genes in mouse spermatozoa (Erkek et al., 2013). These findings endow H3.3 and corresponding modifications prime candidacies for epigenetic inheritance across generations.

Besides of normal cell development, H3.3 is proposed to be a critical epigenetic activator that initiates cell reprogramming to a pluripotency state. Study in somatic cell nuclear transfer (SCNT) embryos suggests that H3.3, but not H3.1/H3.2, is a

maternal “reprogramming factor” that is essential for chromatin reorganization in the donor nucleus, resulting in pluripotency gene reactivation that reprograms a differentiated cell to a ESC-like cell (Jullien et al., 2012; Wen et al., 2014a). In mouse ESCs, H3.3 is suggested to be required for proper establishment of K4me3+K27me3 bivalent chromatin modification, an epigenetic character that is specific to pluripotent stem cells. Several lineage-specific genes are misregulated with reduced levels of K27me3 in H3.3-depleted ESCs, resulting in alteration of the differentiation potential of ESCs (Banaszynski et al., 2013). It consists with the finding that, although knocking out H3.3 specific chaperon HIRA in mouse ESCs gives no observable effect, HIRA-/- mutant causes embryonic lethality at the gastrula stage (Roberts et al., 2002).

We therefore explored the potential role of H3.3 in facilitating cell reprogramming to pluripotency. The well studied transcription factor-induced pluripotent stem cells (iPS) formation method was applied to model general cell reprogramming process. To study the activation process of pluripotency genes, time-series studies on both transcriptome and epigenome of the reprogrammed MEFs (rMEFs) were performed during the first ten days of iPS formation. Our results suggest that global H3.3 replacement initiates at the early stage of MEFs reprogramming, particularly at the promoters and enhancers of ESC signature genes. This H3.3 replacement establishes a pluripotent ESC-like chromatin landscape prior to corresponding gene activation. Besides, promoter H3.3 replacement not only associates with K27me3 reduction at ESC signature genes, but also with re-establishment of bivalent modification at embryonic differentiation genes.

III. Result

1. Global H3.3 replacement initiates right after the induction of MEFs reprogramming.

To understand the potential role of H3.3 in mammalian cell reprogramming, we applied ChIP-seq to identify genomic loci with dynamic H3.3 deposition during transcription factor-induced pluripotent stem cells (iPS) formation (Table 2.1). To this end, mice carrying a single dox-inducible Oct4/Sox2/Klf4/Myc (OSKM) polycistronic 4F2A cassette (tetO-4F2A) (Carey et al., 2010) were crossed with H3.3-HA mice (Wen et al., 2014c). MEFs were then isolated from E13.5 embryos of this tetO-4F2A x H3.3-HA strain and cultured with doxycycline to give rise to germline-transmission competent iPS (data not shown). To focus on early reprogramming events, reprogramming MEFs (rMEFs) were collected at 0, 3, 6 and 10 days after doxycycline induction, then applied to ChIP-seq on HA epitope tag of H3.3 and RNA-seq for gene expression profile (Table 2.2). ESCs and completely reprogrammed iPS (ESC-like cells over 20 days of induction) were used as pluripotent cell controls. The first ten days were considered as early stage of MEFs reprogramming, and time after ten days was considered as late stage of reprogramming (Figure 2.1). By comparing with ChIP-seq result in MEFs without doxycycline induction (day 0), a H3.3 replacement locus was selected as any genomic region in rMEFs (day 3, 6, 10) or iPS with H3.3 ChIP-seq average read count increasing over 0.15, a threshold selected basing on the distribution of H3.3 ChIP-seq read count change at all promoters (Figure 2.2A). (This system can detect that H3.3 replaces H3.1 or H3.2, but is not able to detect when H3.3 replaces H3.3).

Intriguingly, time-series result suggests that H3.3 replacement initiates as early as day 3 after doxycycline induction, the first time point we collected. We found over

one-fourth of promoters (-2kb ~ TSS) with H3.3 replacement in iPS, which is higher than other genetic loci (Figure 2.2B). It suggests that promoter is the most active region for H3.3 replacement during early reprogramming. We found significant H3.3 replacement at 7.2% to 13.8% promoters within ten days of induction, and at 27.4% promoters in completely reprogrammed iPS (Figure 2.2B). Similar percentage of H3.3 replacement was observed within gene body and distal region (over 2kb from transcript), but not at the end of transcipts (downstream) (Figure 2.2B). As a validation of sequencing result, we also examined H3.3 enrichment at promoters of several pluripotency genes by ChIP-qPCR, which showed that H3.3 deposition significantly increases from day 3 to day 6, while total H3 level remains unchanged or only slightly increased (Figure 2.2C). This result suggests that the increase of H3.3 deposition at these loci is a result of H3.3 replacement, not increase of nucleosomal density.

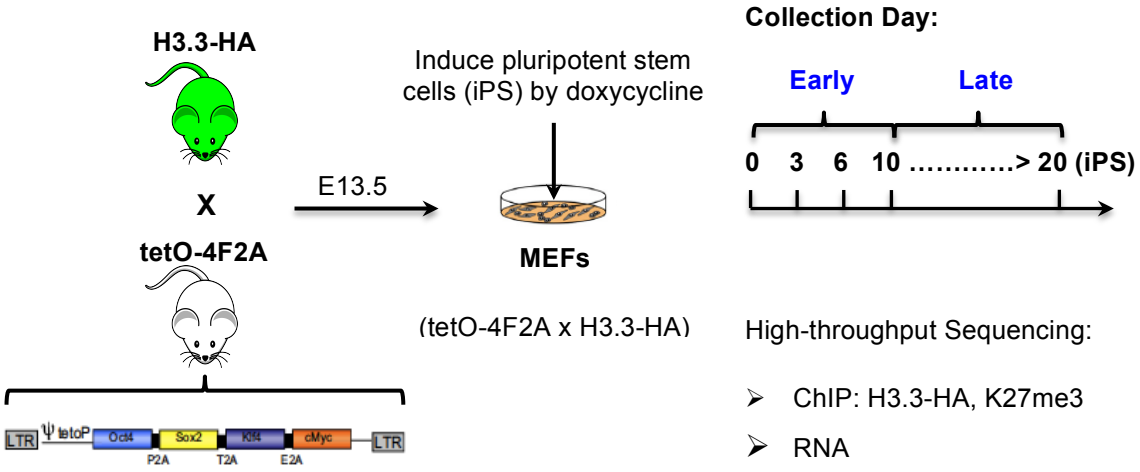


Figure 2.1. Schematic of experimental system and time points of data collection and sequence analyses. (All the mouse work and cell collection were completed by Dr. Duancheng Wen)

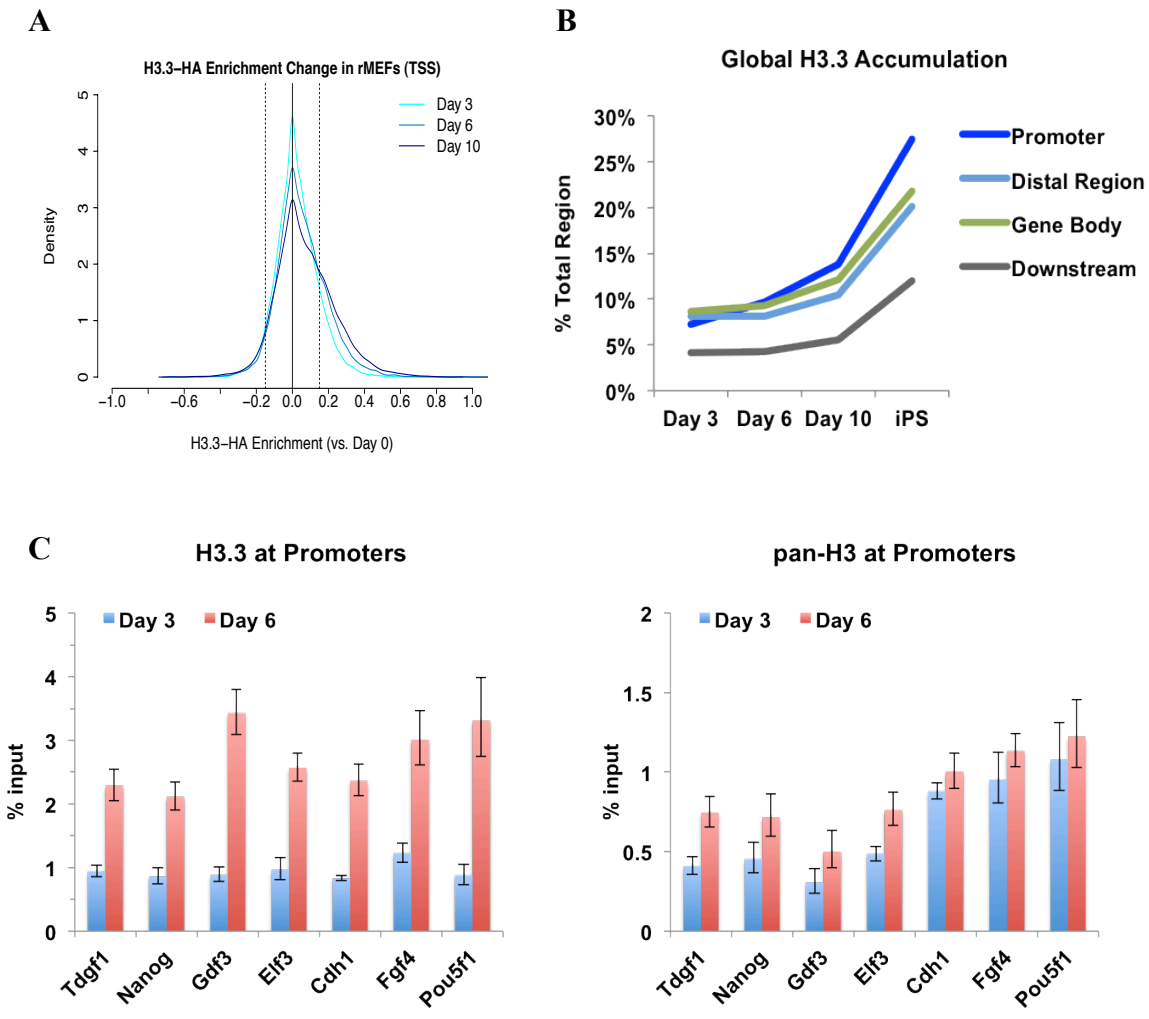


Figure 2.2. H3.3 replacement initiates at promoter and other genomic regions upon the induction of MEFs reprogramming.

(A) Probability distribution of H3.3 enrichment change at all promoters. X-axis, average H3.3 ChIP-seq read count difference between certain time point and MEFs (day 0). Y-axis, density of probability. Light blue, day 3. Blue, day 6. Dark blue, day 10. Dashed line, threshold of H3.3 increase (0.15, replacement) or decrease (-0.15).

(B) Percentage of genomic regions with H3.3 replacement during MEFs reprogramming. H3.3 replacement loci were selected by comparing ChIP-seq signal between certain time point and MEFs (day 0). Dark blue, promoter (-2 kb ~ TSS). Light blue, distal region (over 2 kb from transcript). Green, gene body (including exon and intron). Grey, downstream (TES ~ 2 kb).

(C) ChIP-qPCR on H3.3 (left) and general H3 (right) at selected promoters. (This experiment was completed and analyzed by Dr. Laura A. Banaszynski)

2. Promoter H3.3 replacement establishes a pluripotent ESC-like chromatin landscape prior to gene activation.

To assess the relationship between H3.3 replacement and gene activation during cell reprogramming, we then focused on genes that are *de novo* activated in completely reprogrammed iPS (iPS activated genes). These genes were defined as being completely silenced or expressed at low level in MEFs (day 0) but increasing their expression over 2 folds (log2-based) in iPS (see Methods) (Figure 2.3A). Basing on these criteria, we selected 2,069 iPS activated genes from RNA-seq result for in-depth analyses (Table 2.3). As a control, we randomly selected 2,000 genes that remain silent or low expression in both MEFs and iPS (iPS stable genes) (see Methods) (Figure 2.3A, Table 2.4).

Strikingly, H3.3 replacement was detected at the promoters of 43.1% (891) iPS activated genes within the first ten days of reprogramming (early replacement), and at 27.3% (565) promoters after ten days but in iPS (late replacement) (Figure 2.3B, Table 2.3). Both groups maintained high promoter H3.3 level in iPS (Figure 2.4A). Within the whole reprogramming period, H3.3 deposition did not significantly increase at the rest of 29.6% (613) iPS activated gene promoters (no replacement) (Figure 2.3B, Table 2.3). Within 2,000 iPS stable genes, however, only 22.5% (449) and 18.2% (364) promoters were found with early and late H3.3 replacement (Figure 2.3B, Table 2.4). Conversely, nearly 60% iPS stable gene promoters did not have H3.3 replacement but remained low level of H3.3 throughout MEFs reprogramming (Figure 2.3B, 2.4B, Table 2.4). We also noticed that iPS activated genes had early promoter H3.3 replacement initiated immediately after doxycycline induction (day 3), and reached the iPS-like H3.3 deposition level at day 10 (Figure 2.4A).

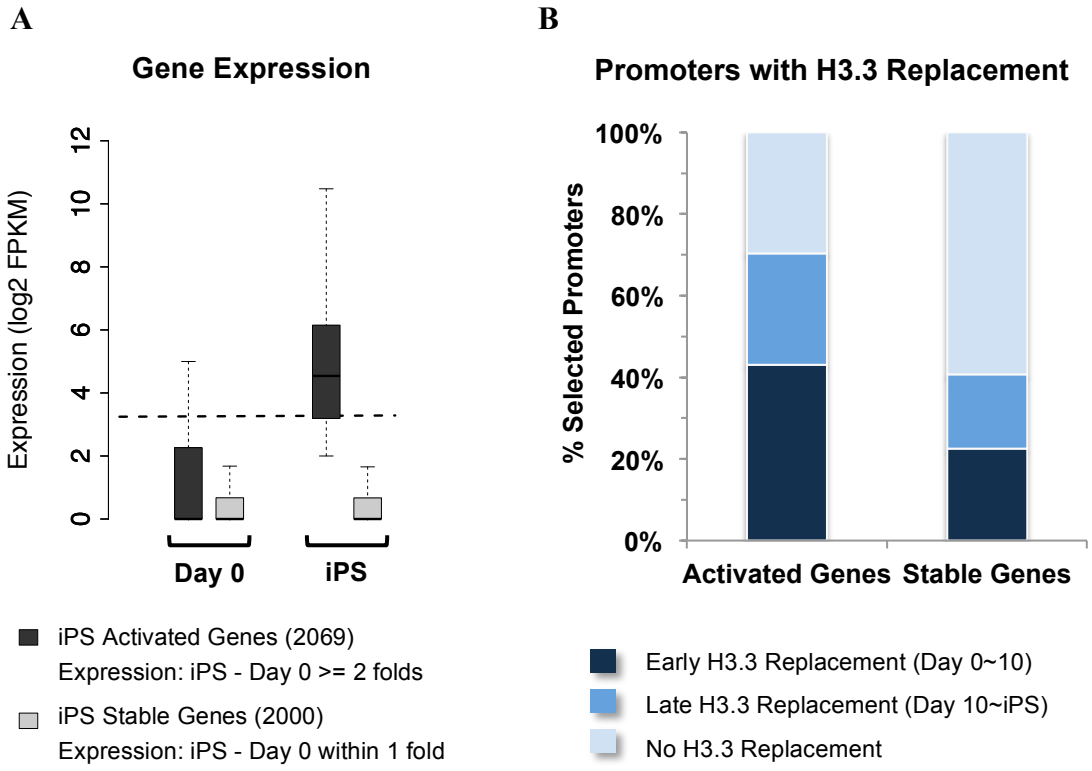


Figure 2.3. iPS activated genes are significantly enriched with H3.3 at promoter.

(A) Expression of 2069 iPS activated genes (black) and 2000 iPS stable genes (grey) in MEFs (day 0) and iPS. Y-axis, gene expression (log2-transformed). Grey dashed line, average of all gene expression at corresponding time point.

(B) Percentage of promoters with H3.3 replacement within the first ten days of reprogramming (Early, dark blue), after ten days but in iPS (Late, blue), or without H3.3 replacement (No, light blue).

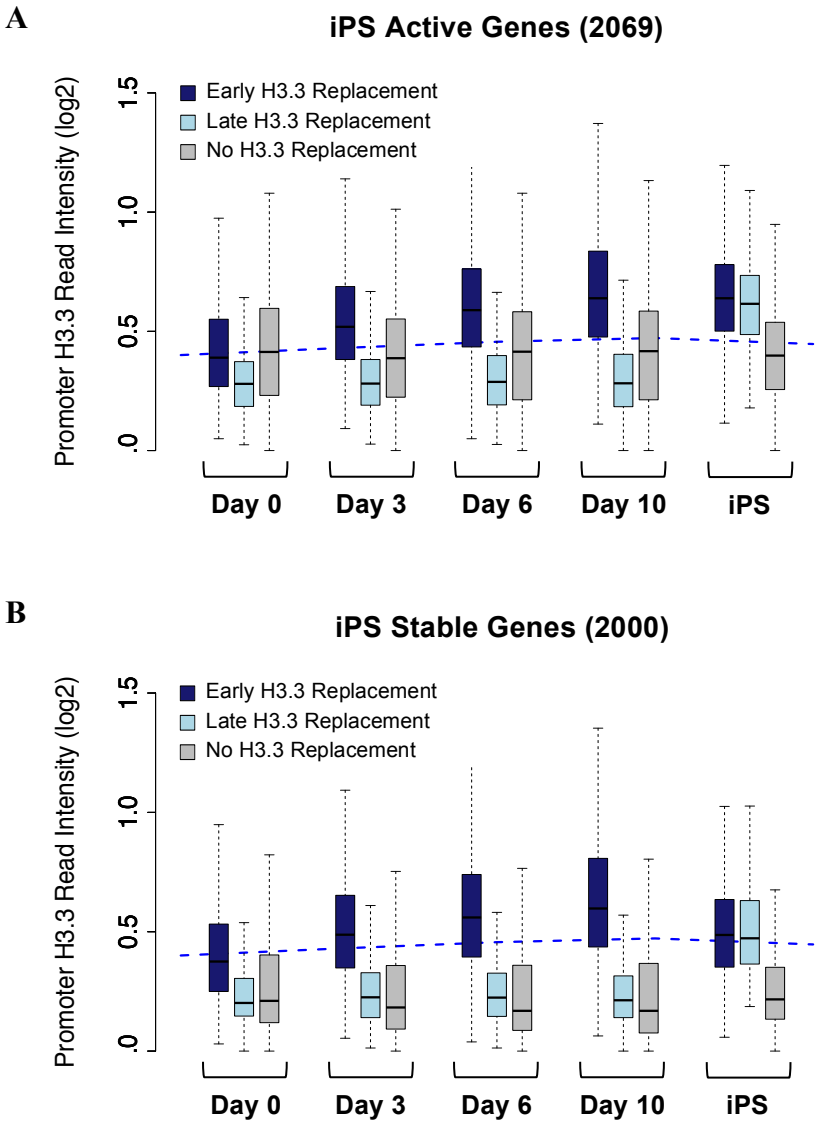


Figure 2.4. Promoter H3.3 replacement during MEFs reprogramming. iPS activated genes are significantly enriched with H3.3 at promoter.

H3.3 enrichment at (A) iPS activated gene promoters and (B) iPS stable gene promoters. Dark blue, early H3.3 replacement. Light blue, late H3.3 replacement. Grey, no H3.3 replacement. Y-axis, average read density within promoter region (-2 kb ~ TSS) with log2-transfer. Blue dashed line, average of H3.3 enrichment at all promoters at corresponding time point.

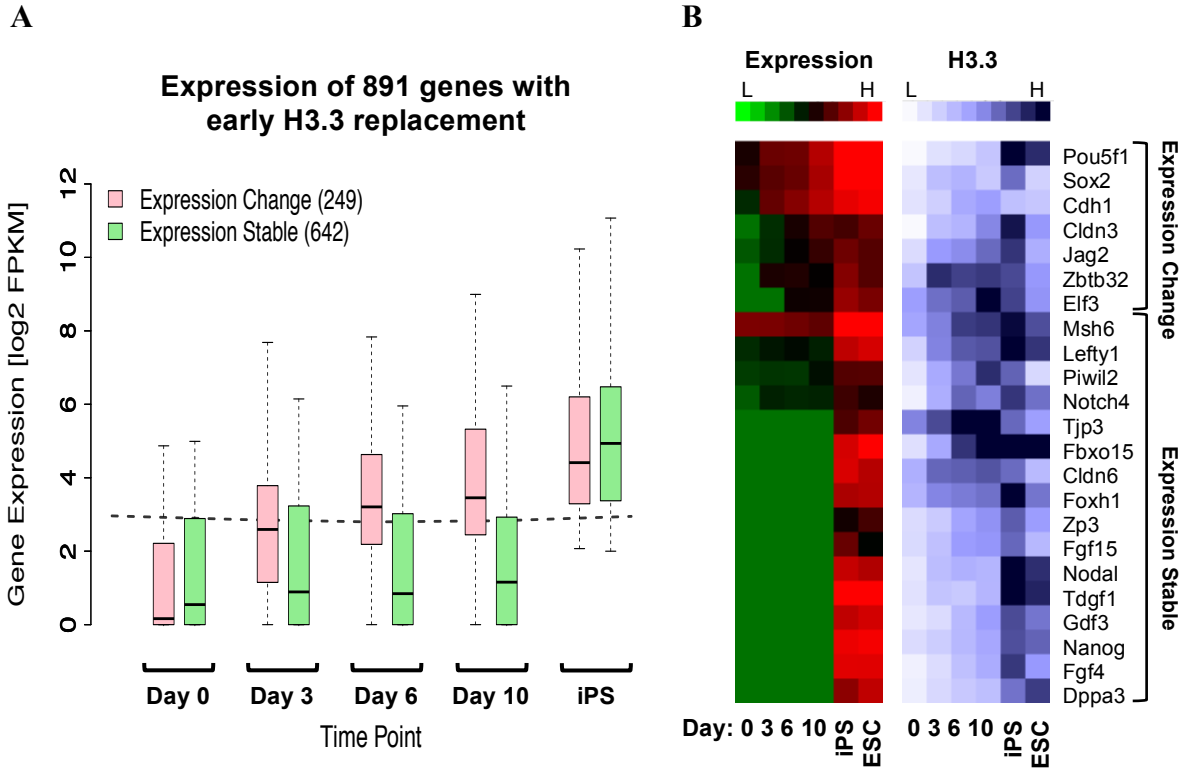
We then studied the relationship between promoter H3.3 replacement and transcription activation. To our surprise, within the 891 iPS activated genes with early H3.3 replacement at promoters, only less than one-third (249) genes increased their expression during the period of H3.3 replacement (Figure 2.5A, B). This result suggests, intriguingly, that the majority of genes undergoing promoter chromatin remodeling do not change their transcription together with H3.3 replacement. Importantly, these genes were expressed later in iPS (Figure 2.5A, B). Gene ontology (GO) analysis revealed that iPS activated genes with either early or late promoter H3.3 replacement mainly regulate ESCs self-renewal and differentiation (Figure 2.5C, Table 2.5), including many key pluripotency genes such as Pou5f1/Oct4, Sox2 and Nanog (Figure 2.5B).

Figure 2.5. Promoter H3.3 replacement establishes a pluripotent chromatin landscape prior to gene activation.

(A) Expression of 891 iPS activated genes with early H3.3 replacement at promoters. Genes are divided by expression patterns during the first ten days of reprogramming, including expression change (249 genes, pink box) and expression stable (642 genes, green box). Y-axis, gene expression (log2-transformed). Grey dashed line, average of all gene expression at corresponding time point.

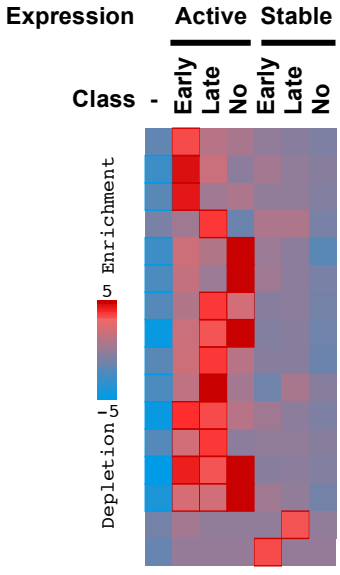
(B) Transcription and H3.3 enrichment profiling for selected iPS activated genes at different time points of MEFs reprogramming. Left, gene expression, red and green indicate relative higher and lower expression, respectively. Right, H3.3 enrichment measured by average read count within promoter region, blue and white indicate relative higher and lower H3.3 enrichment, respectively.

(C) Gene ontology (GO) enrichment using iPAGE for iPS activated genes (Active) and iPS stable genes (Stable). Genes were grouped by promoter H3.3 replacement type as in Figure 2.3B. Enrichment and depletion of certain GO functions (shown on the right) are measured by hypergeometric p-value (log10-transformed). The first column contains control gene genes that do not belong to any of the other 6 groups. Red, over representation. Blue, under representation.



C

Gene Ontology



Early replacement: Day 0 ~ 10
Late replacement: Day 10 ~ iPS
No replacement: Day 0 ~ iPS

Embryonic placenta development, GO:0001892
Stem cell maintenance, GO:0019827
Response to nutrient, GO:0007584
Cell-cell signaling, GO:0007267
DNA metabolic process, GO:0006259
DNA recombination, GO:0006310
Formation of primary germ layer, GO:0001704
Chromosome organization, GO:0051276
Mesoderm development, GO:0007498
Cell fate commitment, GO:0045165
Stem cell differentiation, GO:0048863
Blastocyst formation, GO:0001825
Meiosis, GO:0007126
Sexual reproduction, GO:0019953
Thyroid hormone metabolic process, GO:0042403
Cellular response to abiotic stimulus, GO:0071214

3. Promoter H3.3 replacement associates with K27me3 reduction at ESC signature genes and re-establishment of bivalent modification at embryonic differentiation genes.

It has been suggested that chromatin histone modification is key to transcriptional regulation and cell reprogramming. In particular, K27me3 modification at promoters associates with gene silencing, providing tight control over transcription during both iPS formation and ESC differentiation. Therefore, we examined whether H3.3 replacement associated with gene activation with the adjustment of corresponding promoter K27me3 modification during MEFs reprogramming. rMEFs were collected in the time-series study as described above and applied to ChIP-seq on K27me3 modification at each time point (Figure 2.1).

Among 24,033 unique transcripts in our study, we identified 8,722 promoters with early H3.3 replacement (first ten days) during MEFs reprogramming, and nearly half of them (4,264) had marginally detectable K27me3 modification before doxycycline induction (day 0) (Figure 2.6A, Table 2.6). These low K27me3 promoters were mainly transcribed in both MEFs and iPS (Figure 2.6B) with housekeeping functions in metabolic process and cell cycle (Figure 2.6C, Table 2.7).

The rest of 4,458 promoters with early H3.3 replacement were enriched with K27me3 modificaiton in MEFs (day 0) (Figure 2.6A, Table 2.6). Strickingly, 41.0% of them (1,829/4,458) lost K27me3 during reprogramming, significantly higher than those promoters without H3.3 replacement (23.2%, 1,904/8,192) (Fisher's exact test, p-value < 2.2e-16) (Figure 2.7A). Notably, 85.1% (Class I, 1,557/1,829) promoters progressively lost K27me3 during early reprogramming, with the modification changed together with H3.3 replacement as early as at day 3 (Figure 2.7B, C). These promoters mainly regulate stem cell development genes, including the key

pluripotency genes Pou5f1/Oct4 and Nanog, as well as many pluripotency-associated genes (Figure 2.8, 2.9). This result suggests that K27me3 depletion at pluripotency gene promoters concurrented with H3.3 replacement upon initiation of cell reprogramming.

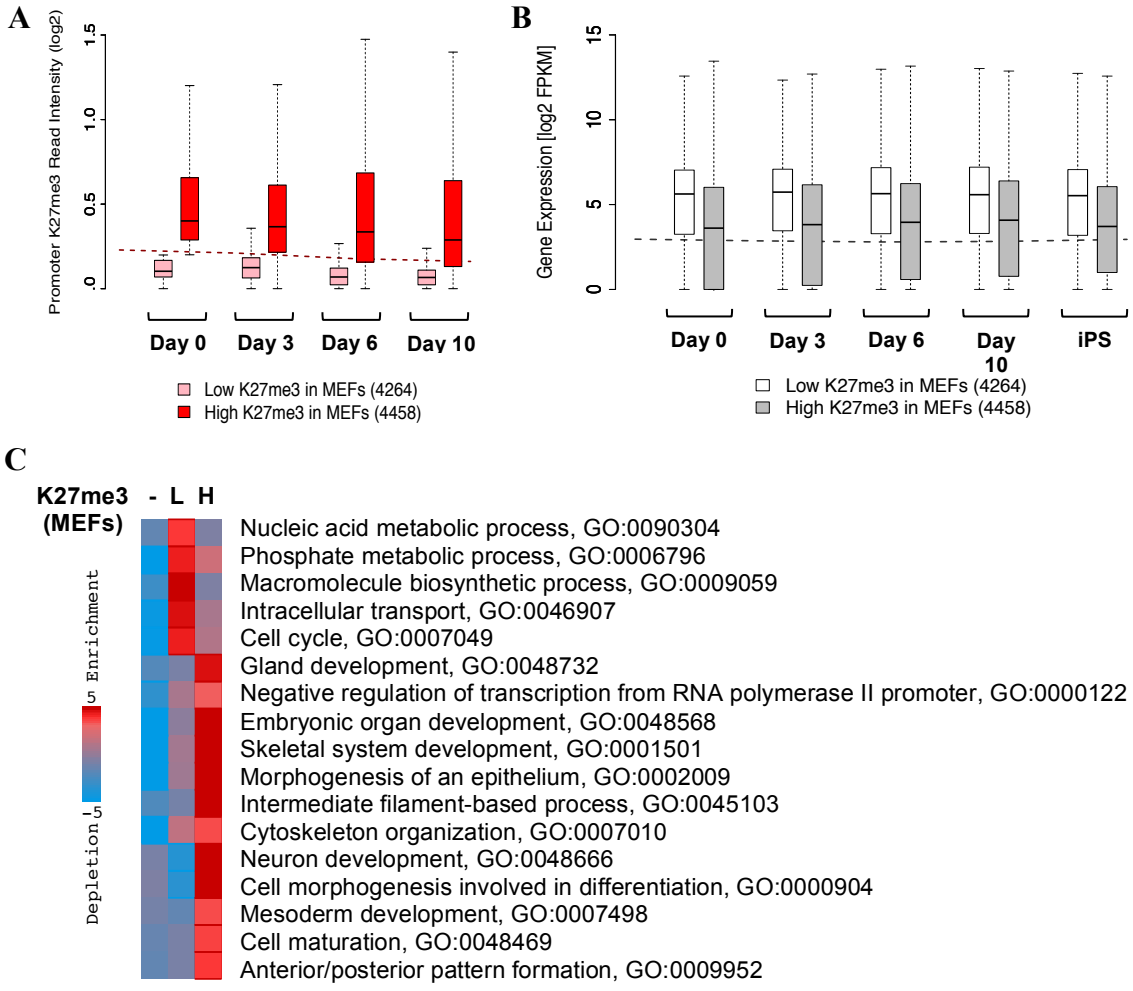


Figure 2.6. Promoters with high K27me3 in MEFs and early H3.3 replacement during reprogramming are enriched for developmental functions.

(A) K27me3 modification change at promoters with early H3.3 replacement. Y-axis, average read density within promoter region (log2-transformed). Red dashed line, average of K27me3 enrichment at all promoters at corresponding time point.

(B) Expression of genes with early H3.3 replacement at promoters. Genes are grouped by K27me3 level as in Figure 2.6A. Y-axis, gene expression (log2-transformed). Grey dashed line, average of all gene expression at corresponding time point.

(C) Gene ontology (GO) enrichment in genes with low promoter K27me3 in MEFs (4264 genes, L) and genes with high promoter K27me3 in MEFs (4458 genes, H).

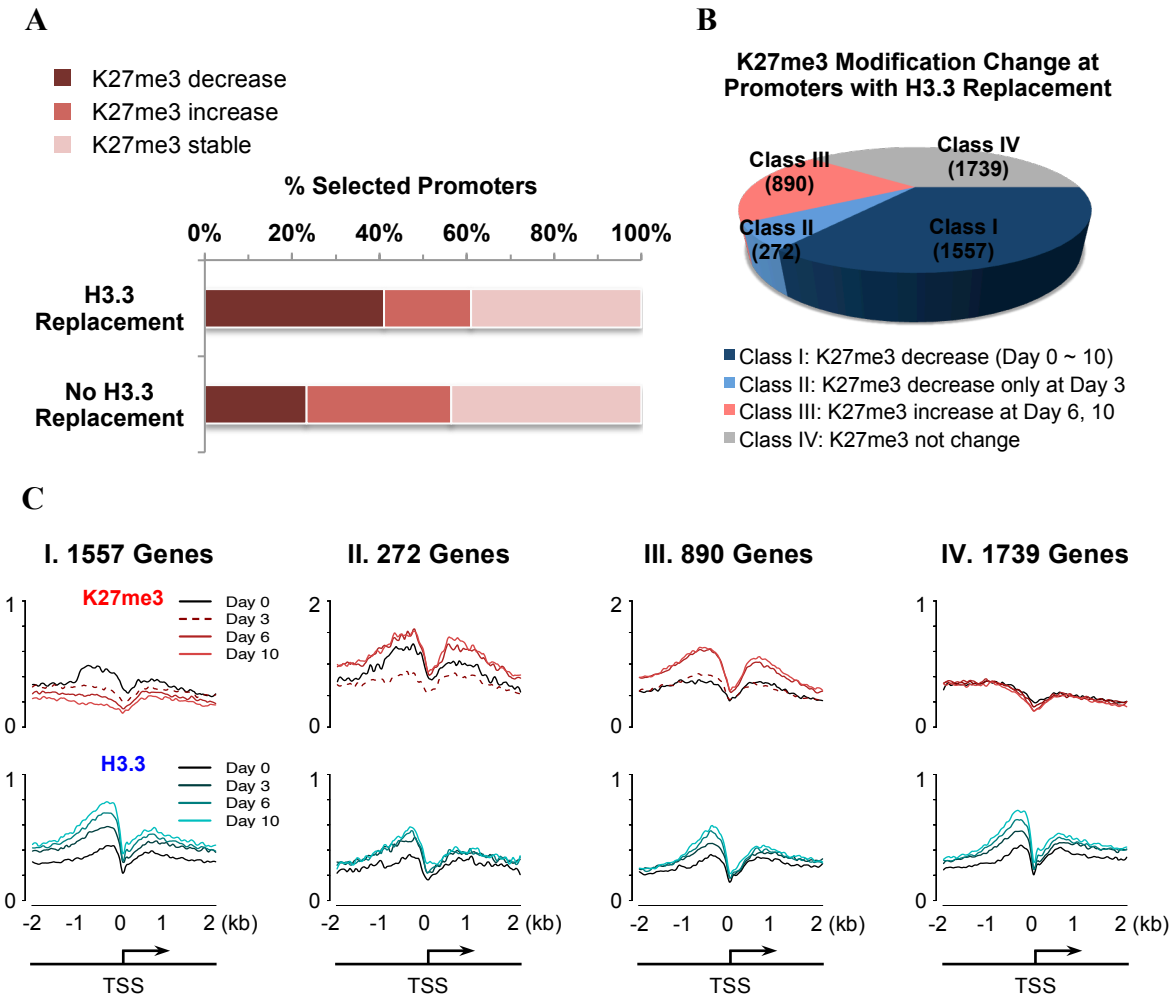


Figure 2.7. Promoter H3.3 replacement associates with the reduction of K27me3 modification.

(A) Percentage of promoters with K27me3 decrease (dark red), increase (red), or stable (pink) during early MEFs reprogramming (the first ten days). Promoters with H3.3 replacement, 4458 genes. Promoters without H3.3 replacement, 8192 genes.

(B) Promoters with high K27me3 modification in MEFs (day 0) and H3.3 replacement in rMEFs are classified by change of K27me3 modification during reprogramming.

(C) K27me3 (top) and H3.3 (bottom) enrichment profiles near the promoters. TSS, transcription start site. Gene grouped as in Figure 2.7B.

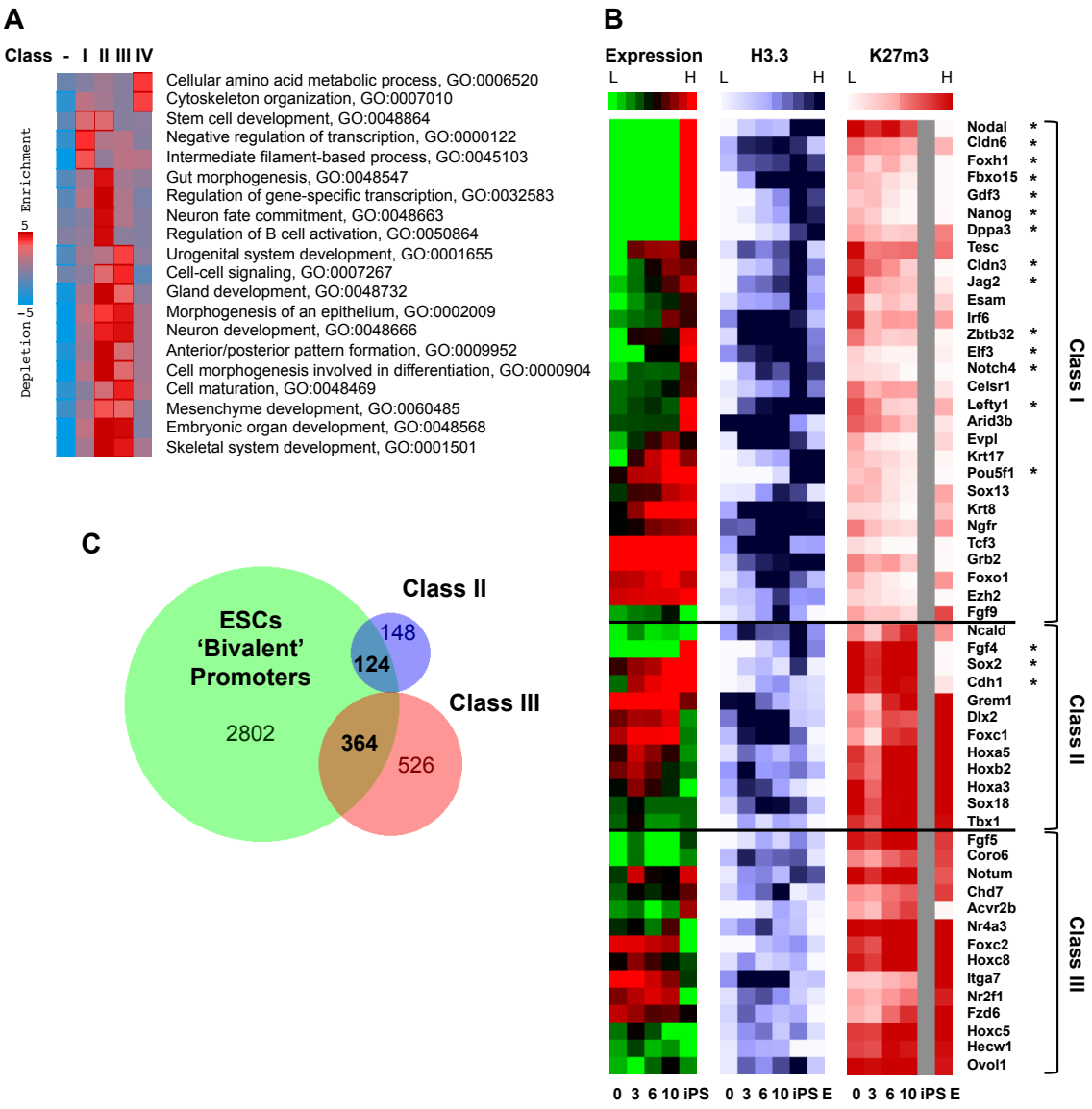


Figure 2.8. Promoters with H3.3 replacement and K27me3 reduction are enriched with ESCs pluripotency- and differentiation-regulation functions.

(A) Gene ontology (GO) enrichment for genes grouped as in Figure 2.7B.

(B) Transcription, H3.3 enrichment, and K27me3 modification profilings for selected promoters during the 10-day period of reprogramming, iPS, and ESCs (E). Left, gene expression, red and green indicate relative higher and lower expression, respectively. Middle, H3.3 enrichment measured by average read count within promoter region, blue and white indicate relative higher and lower H3.3 enrichment, respectively. Right, K27me3 modification measured by average read count within promoter region, red and white indicate relative higher and lower modification, respectively.

(C) Venn diagram among genes with K27me3 increased at promoters during 10-days of induction (Class II, III) and ESC bivalent genes.

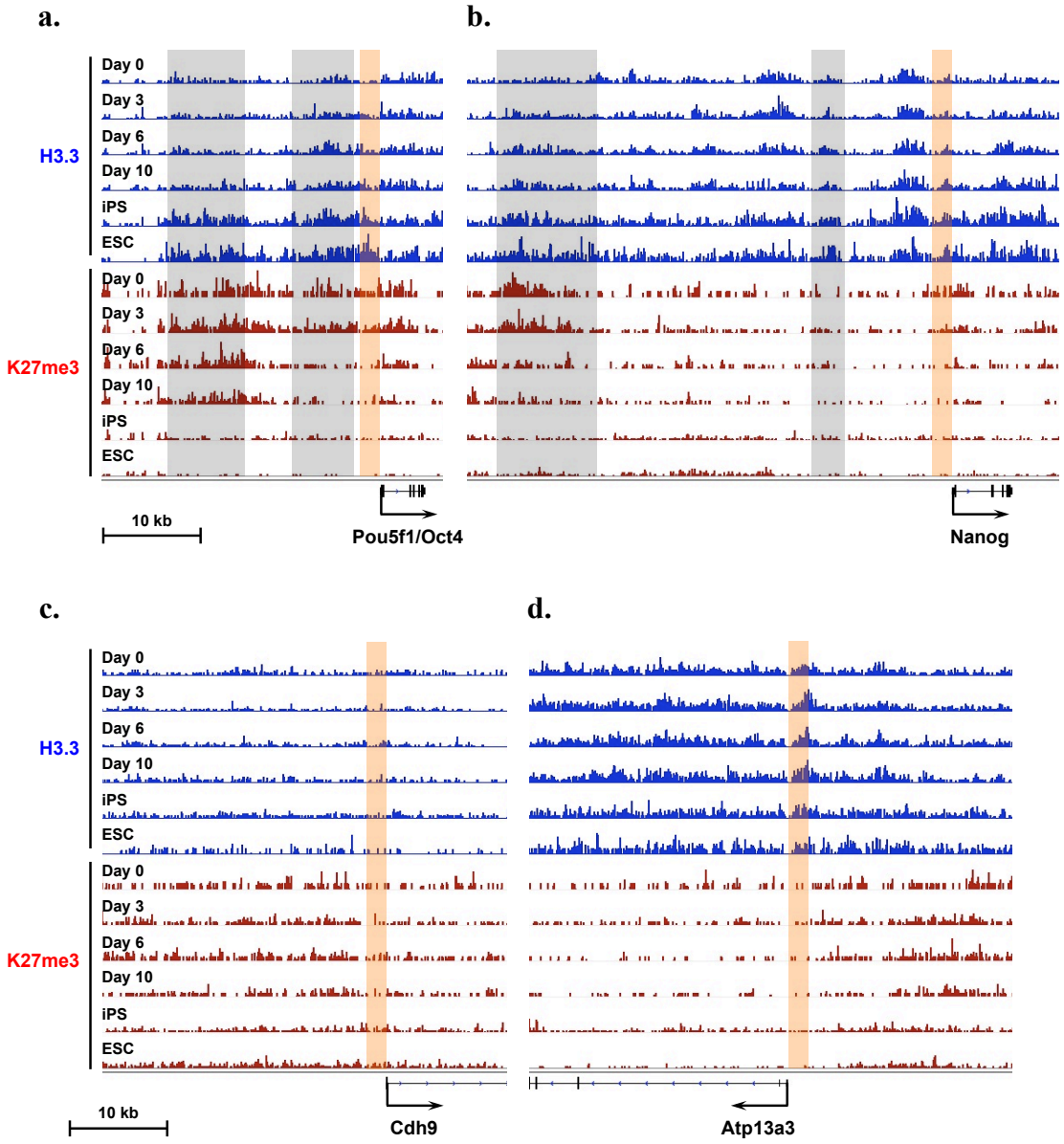


Figure 2.9. H3.3 and K27me3 enrichment at selected promoters. (a) Pou5f1; (b) Nanog; (c) Cdh9; (d) Atp13a3. Blue tracks, H3.3. Red tracks, K27me3. Red region, promoter selected for analysis (-2 kb ~ TSS). Grey region, enhancer. Data range for all tracks, 0 – 0.4.

Intriguingly, we also found many promoters with significant K27me3 upregulation at day 6 and day 10, some lost K27me3 at day 3 (Class II, 272 promoters), but some did not (Class III, 890 promoters) (Figure 2.7B, C, Table 2.6). As Class I promoters, both Class II and III promoters had H3.3 replacement as early as at day 3 and maintained it over the course of reprogramming (Figure 2.7C). However, these promoters mainly regulate embryonic development genes particularly function in mesodermal and epithelial lineages (e.g., Cdh1, Gdf7, Gata6) (Figure 2.8B), making them functionally distinct from Class I promoters or those with early H3.3 replacement but stable K27me3 modification (Class IV, 1,739 promoters) (Figure 2.8A, Table 2.6, 2.8). Moreover, over 40% (488/1,162) of promoters that acquired K27me3 modification after H3.3 replacement (Class II+III) were marked by both K27me3 and K4me3 modifications in ESCs, which were classified as bivalent promoters and considered specific epigenetic signatures in pluripotent stem cells (Figure 2.8C).

4. H3.3 replacement at enhancers reveals transcriptional regulatory hierarchy during early stage of MEFs reprogramming.

Genome-wide studies have shown that in addition to promoters, H3.3 is deposited at enhaners along with recruitment of cell-type specific transcription factors (Goldberg et al., 2010). Within the early stage of MEFs reprogramming (first ten days), we found H3.3 enrichment changed at intergenic regions distal from known transcripts, suggesting the H3.3 replacement at enhancers (Figure 2.2B). To identify MEFs reprogramming associated enhancers, we collected K27ac enriched intergenic loci in both MEFs and ESCs in published ChIP-seq data sets (Creyghton et al., 2010), and assigned loci with cell type specific K27ac enrichment MEF- or ESC-specific

enhancers (Figure 2.10A, Table 2.9). As expected, H3.3 progressively increased at ESC-specific enhancers but decreased at MEF-specific enhancers during iPS formation, and this change was detected as early as at day 3 after induction of reprogramming (Figure 2.10B). Besides, we also observed nearly two-thirds of ESC-specific enhancers undergoing both H3.3 replacement and K27me3 decrease during early reprogramming, significantly higher than those enhancers without H3.3 replacement (Fisher's exact test, p-value < 2.2e-16) (Figure 2.10C).

To unveil the molecular mechanisms involved in MEFs reprogramming associated H3.3 replacement, we applied motif search with HOMER on ESC-specific enhancers with *de novo* H3.3 replacement at each time point of early reprogramming and in completely reprogrammed iPS (Figure 2.11, Table 2.10). Motifs preferentially enriched at these enhancers suggest that the exogenous expression of Oct4, Sox2, and Klf4 potentially activated ESC-specific enhancers right after doxycycline induction (day 3). However, c-Myc potentially associated with H3.3 replacement and enhancer activation slightly later, as the enrichment of c-Myc motif was only identified after day 6 of induction and in iPS (Figure 2.11, left). We also found many stem cell differentiation regulators potentially activated ESC-specific enhancers in reprogrammed iPS (e.g., Yy1, Foxp1, etc.), suggesting their roles in establishing developmental potency to multiple lineages. Conversely, MEF-specific enhancers depleting of H3.3 were enriched with no pluripotency associated transcription factors but mesodermal regulators (Figure 2.11, right).

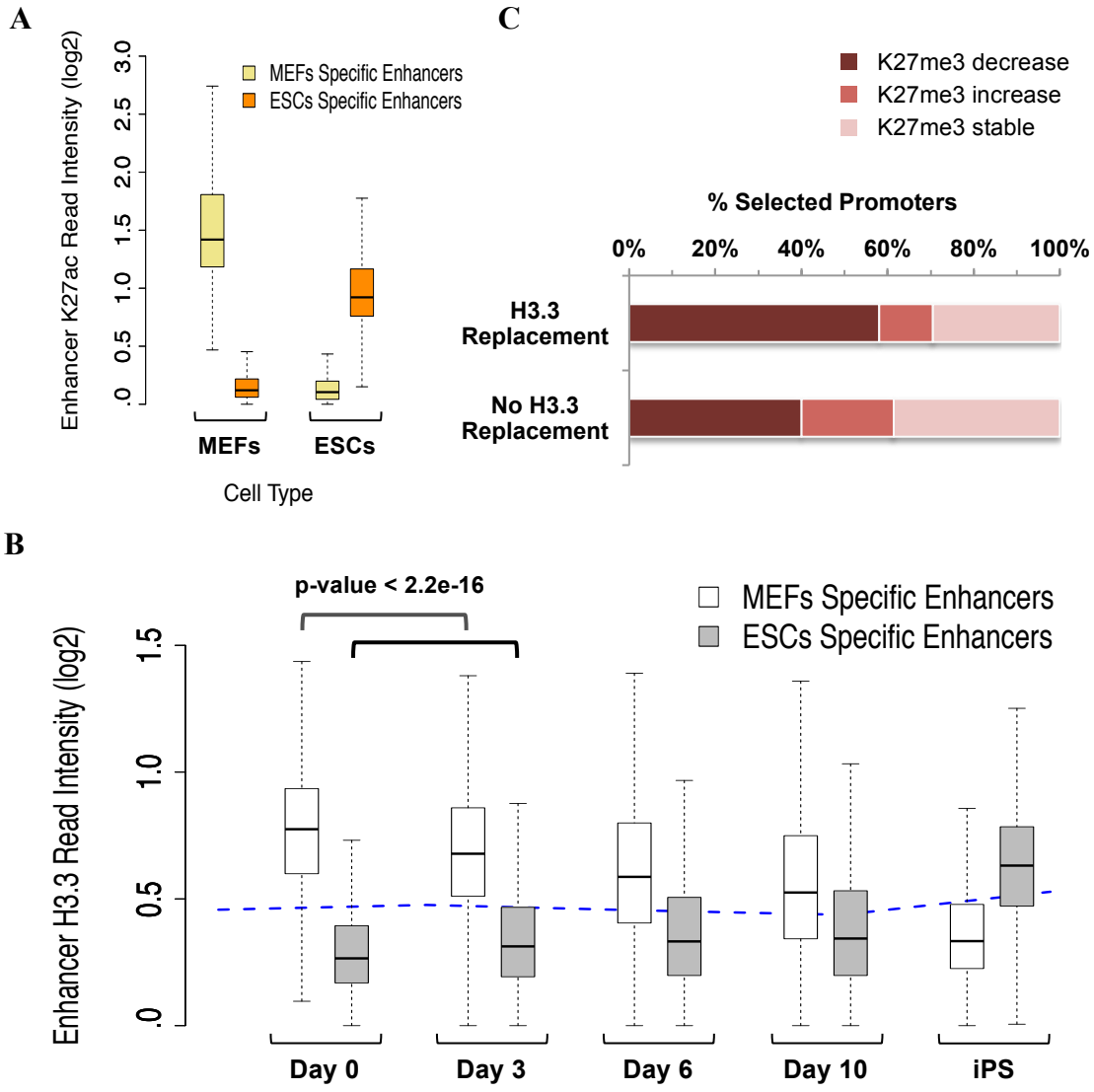


Figure 2.10. ESC-specific enhancers are significantly enriched with H3.3 and reduced for K27me3 during MEFs reprogramming.

(A) K27ac modification at cell type specific enhancers. Y-axis, average read intensity (log2) within selected enhancers. Yellow, MEF-specific enhancer. Orange, ESC-specific enhancer.

(B) H3.3 enrichment at cell type specific enhancers. White, MEF-specific enhancer. Grey, ESC-specific enhancer. Blue dashed line, average of H3.3 enrichment at all enhancers at corresponding time point.

(C) Percentage of ESC specific enhancers with K27me3 decrease (Decrease), increase (Increase), or not change (Stable) within the initial 10 days of reprogramming. Black, ESC enhancer with H3.3 replacement (4935 enhancers). Grey, ESC enhancer without H3.3 replacement (5855 enhancers).

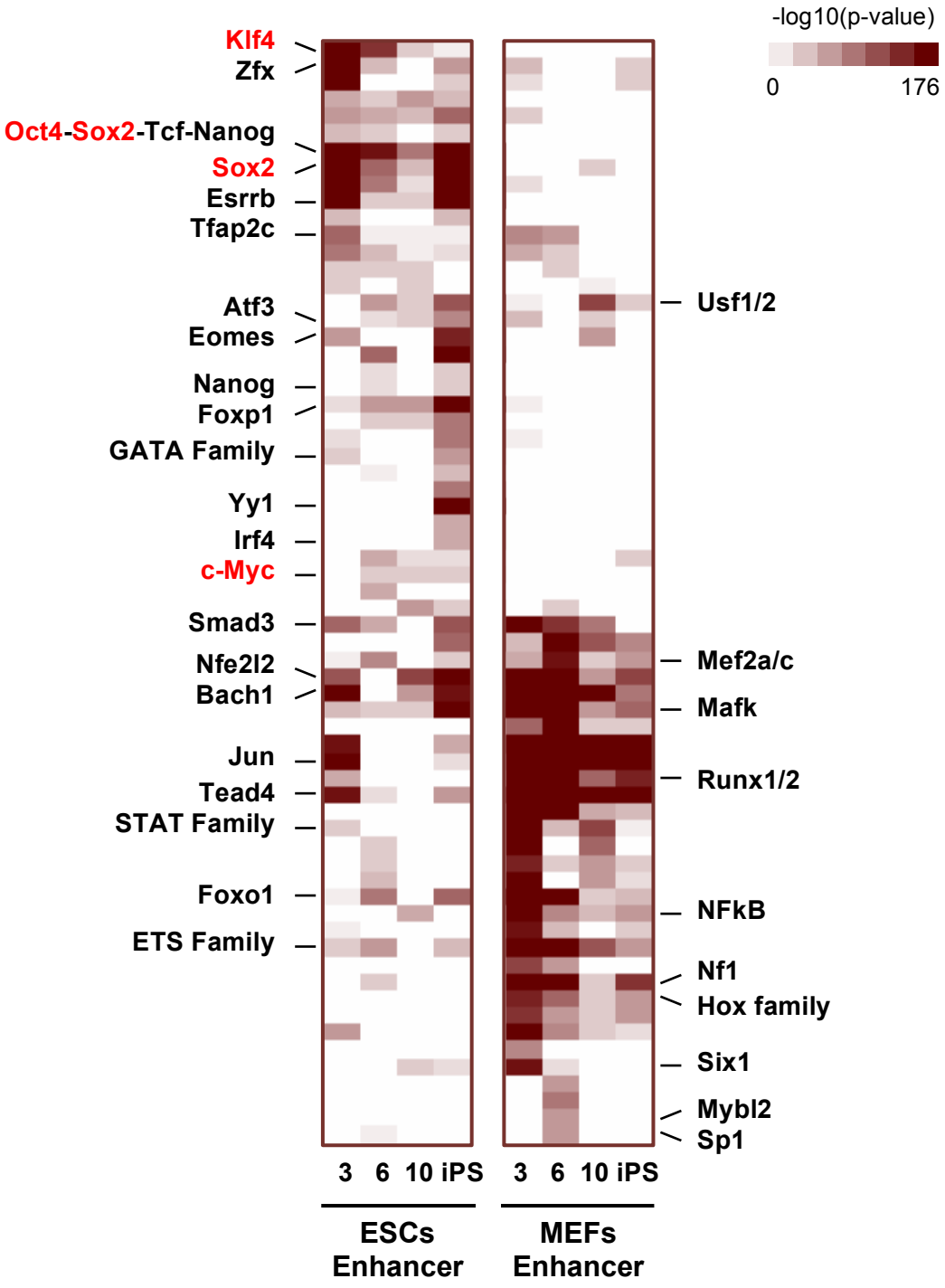


Figure 2.11. Enrichment of transcription factor binding motifs at ESC-specific enhancer (left) and MEF-specific enhancer (right) with H3.3 replacement at different time point. Each row represents a motif, and the corresponding transcription factors are labeled on the side. Red, transcription factors applied for iPS induction.

IV. Discussion

There are several interesting points emerging from our study.

1. H3.3 as an epigenetic regulator that is engaged in cell reprogramming.

Time-series epigenomic study during iPS formation reveals that H3.3 replacement occurs at many pluripotency gene promoters and ESC-specific enhancers immediately after the initiation of MEFs reprogramming. Gene expression profiling at the same time point further suggests that H3.3 replacement takes place prior to corresponding gene activation. These results indicate that rather than simply marking regions of active transcription, *de novo* H3.3 deposition at the transcription regulatory regions of repressed genes may facilitate gene acitivation through the establishment of a pluripotent ESC-like chromatin landscape.

2. H3.3 replacement associates with reduction of K27me3 modification at transcription regulatory loci and induction of pluripotency gene expression.

Our epigenomic study during early and late stages of MEFs reprogramming reveals a close association between H3.3 replacement and K27me3 reduction at both promoters and enhancers. It will be important to evaluate whether H3.3 turnover at these transcription regulatory loci directly induces the decrease of H27me3 modification. Therefore, we hypothesize that through the replacement of chromatin H3.3 with newly synthesized H3.3, K27me3 modification could be removed from chromatin, which turns the H3.3 replacement locus from transcription repressive to permissive status (Figure 2.12). This K27me3 demethylase-independent model suggests a mechanism to remove repressive histone modifications at transcription regulatory loci for pluripotency gene reactivation.

However, we could not exclude the possibility that demethylases function at the H3.3 replacement loci to erase K27me3 modification. It is therefore necessary to evaluate H3.3 replacement-associated K27me3 modification change with the inhibition of K27 demethylases, for example, inhibiting JMJD3 and UTX with GSK-J4 (Figure 2.12) (Kruidenier et al., 2012). We propose that after efficient inhibition on K27 demethylases, K27me3 modification at transcription regulatory loci can still decrease to certain level with H3.3 replacement. Besides, we would confirm H3.3 replacement effect on K27me3 modification change with Hira knockout (Hira-/-) MEFs, which lack H3.3 deposition at transcription regulatory loci (Figure 2.12). We propose that Hira-/- MEFs with K27 demethylases inhibitor treatment can not be effectly reprogrammed, as the pluripotency genes remain repressed with stable K27me3 modification. This defect can be rescued by over expressing H3.3 but not H3.1 or H3.2.

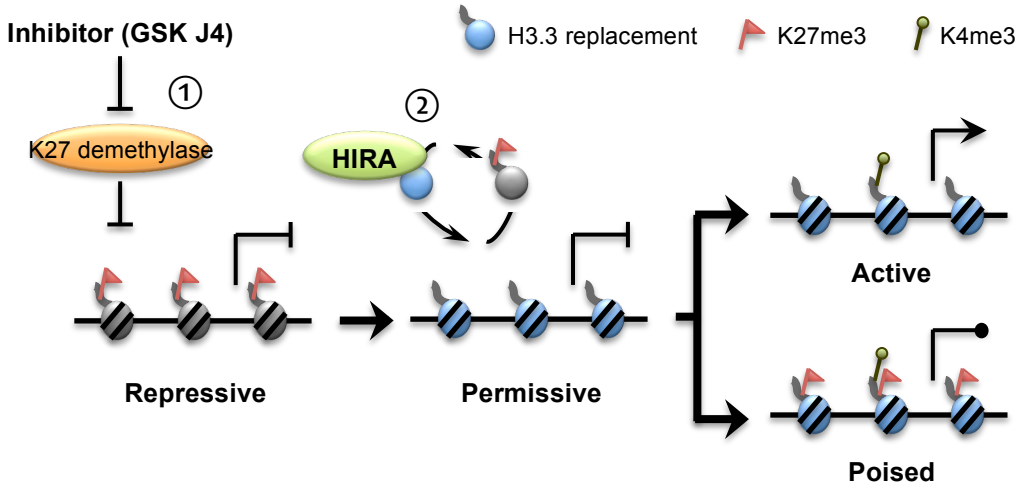


Figure 2.12. Model of H3.3 replacement associated K27me3 modification change and gene activation. The promoter of a silenced gene has repressive modifications on histone H3 (i.e. K27me3, red flag). Newly produced H3.3 (unmodified H3.3) replaces original chromatin histone H3 at this promoter. This replacement removes original repressive modifications and turns the promoter chromatin into a permissive state, which can be transcriptionally activated (i.e. K4me3, green ball) or poised (i.e. K4me3+K27me3) with different histone modifications in reprogrammed cells.

3. H3.3 replacement associates with the establishment of ESC-like chromatin landscape, including bivalent modifications at developmental gene promoters.

Beside of H3.3 replacement associated K27me3 reduction at ESC signature genes, we also found that K27me3 modification was re-established after H3.3 replacement at many developmental gene promoters. Many of these promoters carry K4me3+K27me3 bivalent modification in ESCs and are responsible for ESC differentiation to a broad range of cell types. This result implies that H3.3 replacement associates with proper establishment of bivalency in reprogrammed iPS. This hypothesis is supported by the recent study in H3.3-depleted mouse ESCs, which have K27me3 level reduced at several bivalent gene promoters upon removing of H3.3 protein (Banaszynski et al., 2013). It has been suggested that H3.3 replacement facilitates a dynamic chromatin environment that allows for PRC2 binding and methylation at K27. Therefore, H3.3 replacement during the early stage of MEFs reprogramming potentially directs the establishment of bivalency at developmental gene promoters. It will be interesting to check whether H3.3-depleted MEFs can still go through reprogramming process to generate pluripotent iPS, and whether promoter bivalency can be properly established in reprogrammed cells.

In summary, identification of ESC-like H3.3 enrichment in unipotent SSCs (Chapter one) raises an intriguing question that, besides of K4me3+K27me3 bivalent histone modification, H3.3 could be another epigenetic regulator that pre-deposit pluripotent developmental potential in unipotent SSCs. Our time-series study on global H3.3 replacement during iPS formation further suggests that H3.3 is a novel pluripotency activator that is involved at the initiation of cell reprogramming. In the future, it will be interesting to investigate whether SSC reversion can be improved

through overexpression of H3.3, as transcription factors do for iPS formation.

REFERENCES

Agger, K., Cloos, P.A.C., Christensen, J., Pasini, D., Rose, S., Rappsilber, J., Issaeva, I., Canaani, E., Salcini, A.E., and Helin, K. (2007). UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. Nature *449*, 731–734.

Ahmad, K., and Henikoff, S. (2002). The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. Mol Cell *9*, 1191–1200.

Arnold, K., Sarkar, A., Yram, M.A., Polo, J.M., Bronson, R., Sengupta, S., Seandel, M., Geijsen, N., and Hochedlinger, K. (2011). Sox2+ Adult Stem and Progenitor Cells Are Important for Tissue Regeneration and Survival of Mice. Stem Cell *9*, 317–329.

Baba, S., Heike, T., Umeda, K., Iwasa, T., Kaichi, S., Hiraumi, Y., Doi, H., Yoshimoto, M., Kanatsu-Shinohara, M., Shinohara, T., et al. (2007). Generation of cardiac and endothelial cells from neonatal mouse testis-derived multipotent germline stem cells. Stem Cells *25*, 1375–1383.

Banaszynski, L.A., Allis, C.D., and Lewis, P.W. (2010). Histone variants in metazoan development. Dev Cell *19*, 662–674.

Banaszynski, L.A., Wen, D., Dewell, S., Whitcomb, S.J., Lin, M., Diaz, N., Elsässer, S.J., Chapgier, A., Goldberg, A.D., Canaani, E., et al. (2013). Hira-Dependent Histone H3.3 Deposition Facilitates PRC2 Recruitment at Developmental Loci in ES Cells. Cell *155*, 107–120.

Barak, O., Lazzaro, M.A., Lane, W.S., Speicher, D.W., Picketts, D.J., and Shiekhattar, R. (2003). Isolation of human NURF: a regulator of Engrailed gene expression. Embo J *22*, 6089–6100.

Barroca, V., Lassalle, B., Coureuil, M., Louis, J.P., Le Page, F., Testart, J., Allemand, I., Riou, L., and Fouchet, P. (2009). Mouse differentiating spermatogonia can generate germinal stem cells in vivo. Nat Cell Biol *11*, 190–196.

Bastos, H., Lassalle, B., Chicheportiche, A., Riou, L., Testart, J., Allemand, I., and Fouchet, P. (2005). Flow cytometric characterization of viable meiotic and postmeiotic cells by Hoechst 33342 in mouse spermatogenesis. Cytometry *65A*, 40–49.

Baudat, F., Buard, J., Grey, C., Fledel-Alon, A., Ober, C., Przeworski, M., Coop, G., and de Massy, B. (2010). PRDM9 Is a Major Determinant of Meiotic Recombination Hotspots in Humans and Mice. Science *327*, 836–840.

Bernstein, B., Mikkelsen, T., Xie, X., Kamal, M., Huebert, D., Cuff, J., Fry, B.,

Meissner, A., Wernig, M., and Plath, K. (2006). A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. Cell *125*, 315–326.

Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J., Gingeras, T.R., et al. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. Cell *120*, 169–181.

Borde, V., Robine, N., Lin, W., Bonfils, S., Géli, V., and Nicolas, A. (2009). Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites. Embo J *28*, 99–111.

Boulanger, C.A., Mack, D.L., Booth, B.W., and Smith, G.H. (2007). Interaction with the mammary microenvironment redirects spermatogenic cell fate in vivo. Proc Natl Acad Sci USA *104*, 3871–3876.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947–956.

Brinster, R.L., and Avarbock, M.R. (1994). Germline transmission of donor haplotype following spermatogonial transplantation. Proc Natl Acad Sci USA *91*, 11303–11307.

Brinster, R.L., and Zimmermann, J.W. (1994). Spermatogenesis following male germ-cell transplantation. Proc Natl Acad Sci USA *91*, 11298–11302.

Brykczynska, U., Hisano, M., Erkek, S., Ramos, L., Oakeley, E.J., Roloff, T.C., Beisel, C., Schübeler, D., Stadler, M.B., and Peters, A.H.F.M. (2010). Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. Nat Struct Mol Biol *17*, 679–687.

Buaas, F.W., Kirsh, A.L., Sharma, M., McLean, D.J., Morris, J.L., Griswold, M.D., De Rooij, D.G., and Braun, R.E. (2004). Plzf is required in adult male germ cells for stem cell self-renewal. Nature Publishing Group *36*, 647–652.

Buganim, Y., Faddah, D.A., Cheng, A.W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S.L., van Oudenaarden, A., and Jaenisch, R. (2012). Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell *150*, 1209–1222.

Bush, K.M., Yuen, B.T., Barrilleaux, B.L., Riggs, J.W., O'Geen, H., Cotterman, R.F., and Knoepfler, P.S. (2013). Endogenous mammalian histone H3.3 exhibits chromatin-related functions during development. Epigenetics Chromatin *6*, 7.

Carey, B.W., Markoulaki, S., Beard, C., Hanna, J., and Jaenisch, R. (2010). Single-gene transgenic mouse strains for reprogramming adult somatic cells. Nat Methods *7*, 56–59.

Carmell, M.A., Girard, A., van de Kant, H.J.G., Bourc'his, D., Bestor, T.H., De Rooij, D.G., and Hannon, G.J. (2007). MIWI2 Is Essential for Spermatogenesis and Repression of Transposons in the Mouse Male Germline. Dev Cell *12*, 503–514.

Carrozza, M., Li, B., Florens, L., Suganuma, T., Swanson, S., Lee, K., Shia, W., Anderson, S., Yates, J., Washburn, M., et al. (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell *123*, 581–592.

Chen, C., Ouyang, W., Grigura, V., Zhou, Q., Carnes, K., Lim, H., Zhao, G.-Q., Arber, S., Kurpios, N., Murphy, T.L., et al. (2005). ERM is required for transcriptional control of the spermatogonial stem cell niche. Nature *436*, 1030–1034.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell *133*, 1106–1117.

Christensen, J., Agger, K., Cloos, P.A.C., Pasini, D., Rose, S., (null), Rappsilber, J., Hansen, K.H., Salcini, A.E., and Helin, K. (2007). RBP2 belongs to a family of demethylases, specific for tri- and dimethylated lysine 4 on histone 3. Cell *128*, 1063–1076.

Costoya, J.A., Hobbs, R.M., Barna, M., Cattoretti, G., Manova, K., Sukhwani, M., Orwig, K.E., Wolgemuth, D.J., and Pandolfi, P.P. (2004). Essential role of Plzf in maintenance of spermatogonial stem cells. Nature Publishing Group *36*, 653–659.

Couldrey, C., Carlton, M.B., Nolan, P.M., Colledge, W.H., and Evans, M.J. (1999). A retroviral gene trap insertion into the histone 3.3A gene causes partial neonatal lethality, stunted growth, neuromuscular deficits and male sub-fertility in transgenic mice. Hum Mol Genet *8*, 2489–2495.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci USA *107*, 21931–21936.

Dann, C.T., Alvarado, A.L., Molyneux, L.A., Denard, B.S., Garbers, D.L., and Porteus, M.H. (2008). Spermatogonial stem cell self-renewal requires OCT4, a factor downregulated during retinoic acid-induced differentiation. Stem Cells *26*, 2928–2937.

Davis, T.L., Trasler, J.M., Moss, S.B., Yang, G.J., and Bartolomei, M.S. (1999). Acquisition of the H19 methylation imprint occurs differentially on the parental alleles during spermatogenesis. Genomics *58*, 18–28.

de Hoon, M.J.L., Imoto, S., Nolan, J., and Miyano, S. (2004). Open source clustering software. Bioinformatics *20*, 1453–1454.

de Rooij, D.G., and Russell, L.D. (2000). All you wanted to know about spermatogonia but were afraid to ask. J Androl *21*, 776–798.

Drané, P., Ouararhni, K., Depaux, A., Shuaib, M., and Hamiche, A. (2010). The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3. Genes Dev *24*, 1253–1265.

Erkek, S., Hisano, M., Liang, C.-Y., Gill, M., Murr, R., Dieker, J., Schübeler, D., van der Vlag, J., Stadler, M.B., and Peters, A.H.F.M. (2013). Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. Nat Struct Mol Biol *20*, 868–875.

Falender, A.E., Freiman, R.N., Geles, K.G., Lo, K.C., Hwang, K., Lamb, D.J., Morris, P.L., Tjian, R., and Richards, J.S. (2005). Maintenance of spermatogenesis requires TAF4b, a gonad-specific subunit of TFIID. Genes Dev *19*, 794–803.

Garcia, B.A., Thomas, C.E., Kelleher, N.L., and Mizzen, C.A. (2008). Tissue-Specific Expression and Post-Translational Modification of Histone H3 Variants. J Proteome Res 13.

Gatewood, J.M., Cook, G.R., Balhorn, R., Bradbury, E.M., and Schmid, C.W. (1987). Sequence-specific packaging of DNA in human sperm chromatin. Science *236*, 962–964.

Giannopoulou, E.G., and Elemento, O. (2011). An integrated ChIP-seq analysis platform with customizable workflows. BMC Bioinformatics *12*, 277.

Ginsburg, M., Snow, M.H., and McLaren, A. (1990). Primordial germ cells in the mouse embryo during gastrulation. Development *110*, 521–528.

Glaser, T., Opitz, T., Kischlat, T., Konang, R., Sasse, P., Fleischmann, B.K., Engel, W., Nayernia, K., and Brustle, O. (2008). Adult germ line stem cells as a source of functional neurons and glia. Stem Cells *26*, 2434–2443.

Godmann, M., (null), (null), Di Sauro, A., (null), (null), and Kimmins, S. (2007). Dynamic regulation of histone h3 methylation at lysine 4 in mammalian spermatogenesis. Biol Reprod *77*, 754–764.

Goldberg, A., Allis, C., and Bernstein, E. (2007). Epigenetics: A Landscape Takes Shape. Cell *128*, 635–638.

Goldberg, A.D., Banaszynski, L.A., Noh, K.-M., Lewis, P.W., Elsaesser, S.J., Stadler, S., Dewell, S., Law, M., Guo, X., Li, X., et al. (2010). Distinct Factors Control Histone Variant H3.3 Localization at Specific Genomic Regions. Cell *140*, 678–691.

Goodarzi, H., Elemento, O., and Tavazoie, S. (2009). Revealing Global Regulatory Perturbations across Human Cancers. Mol Cell *36*, 900–911.

Guan, K., Nayernia, K., Maier, L.S., Wagner, S., Dressel, R., Lee, J.H., Nolte, J., Wolf, F., Li, M., Engel, W., et al. (2006). Pluripotency of spermatogonial stem cells from adult mouse testis. Nature *440*, 1199–1203.

Hake, S.B., and Allis, C.D. (2006). Histone H3 variants and their potential role in indexing mammalian genomes: the "H3 barcode hypothesis." Proc Natl Acad Sci USA *103*, 6428–6435.

Hammoud, S.S., Low, D.H.P., Yi, C., Carrell, D.T., Guccione, E., and Cairns, B.R. (2014). Chromatin and Transcription Transitions of Mammalian Adult Germline Stem Cells and Spermatogenesis. Stem Cell 1–34.

Hammoud, S.S., Nix, D.A., Zhang, H., Purwar, J., Carrell, D.T., and Cairns, B.R. (2009). Distinctive chromatin in human sperm packages genes for embryo development. Nature *460*, 473–478.

Hamra, F.K., Chapman, K.M., Nguyen, D.M., Williams-Stephens, A.A., Hammer, R.E., and Garbers, D.L. (2005). Self renewal, expansion, and transfection of rat spermatogonial stem cells in culture. Proc Natl Acad Sci USA *102*, 17430–17435.

Hawkins, R.D., Hon, G.C., Lee, L.K., Ngo, Q., Lister, R., Pelizzola, M., Edsall, L.E., Kuan, S., Luu, Y., Klugman, S., et al. (2010). Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. Cell Stem Cell *6*, 479–491.

He, J., Kallin, E.M., Tsukada, Y.-I., and Zhang, Y. (2008). The H3K36 demethylase Jhdm1b/Kdm2b regulates cell proliferation and senescence through p15(Ink4b). Nat Struct Mol Biol *15*, 1169–1175.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576–589.

Heng, J.-C.D., Feng, B., Han, J., Jiang, J., Kraus, P., Ng, J.-H., Orlov, Y.L., Huss, M., Yang, L., Lufkin, T., et al. (2010). The nuclear receptor Nr5a2 can replace Oct4 in the reprogramming of murine somatic cells to pluripotent cells. Cell Stem Cell *6*, 167–174.

Huangfu, D., Maehr, R., Guo, W., Eijkelenboom, A., Snitow, M., Chen, A.E., and Melton, D.A. (2008). Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. Nat Biotechnol *26*, 795–797.

Imamura, M., Miura, K., Iwabuchi, K., Ichisaka, T., Nakagawa, M., Lee, J., Kanatsu-Shinohara, M., Shinohara, T., and Yamanaka, S. (2006). Transcriptional repression and DNA hypermethyl in SSC of a subset of ESC-specific genes. BMC Dev Biol *6*, 34.

Ishii, K., Kanatsu-Shinohara, M., Toyokuni, S., and Shinohara, T. (2012). FGF2 mediates mouse spermatogonial stem cell self-renewal via upregulation of Etv5 and

Bcl6b through MAP2K1 activation. Development *139*, 1734–1743.

Iwase, S., Lan, F., Bayliss, P., la Torre-Ubieta, de, L., Huarte, M., Qi, H.H., Whetstine, J.R., Bonni, A., Roberts, T.M., and Shi, Y. (2007). The X-linked mental retardation gene SMCX/JARID1C defines a family of histone H3 lysine 4 demethylases. Cell *128*, 1077–1088.

Izadyar, F., Pau, F., Marh, J., Slepko, N., Wang, T., Gonzalez, R., Ramos, T., Howerton, K., Sayre, C., and Silva, F. (2008). Generation of multipotent cell lines from a distinct population of male germ line stem cells. Reproduction *135*, 771–784.

Jaenisch, R., and Young, R. (2008). Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. Cell *132*, 567–582.

Jin, Y., Rodriguez, A.M., Stanton, J.D., Kitazono, A.A., and Wyrick, J.J. (2007). Simultaneous mutation of methylated lysine residues in histone H3 causes enhanced gene silencing, cell cycle defects, and cell lethality in Saccharomyces cerevisiae. Mol Cell Biol *27*, 6832–6841.

Jullien, J., Astrand, C., Szenker, E., Garrett, N., Almouzni, G., and Gurdon, J.B. (2012). HIRA dependent H3.3 deposition is required for transcriptional reprogramming following nuclear transfer to Xenopus oocytes. Epigenetics Chromatin *5*, 17.

Kanatsu-Shinohara, M., and Shinohara, T. (2013). Spermatogonial Stem Cell Self-Renewal and Development. Annu Rev Cell Dev Biol *29*, 163–187.

Kanatsu-Shinohara, M., Inoue, K., Lee, J., Yoshimoto, M., Ogonuki, N., Miki, H., Baba, S., Kato, T., Kazuki, Y., Toyokuni, S., et al. (2004). Generation of pluripotent stem cells from neonatal mouse testis. Cell *119*, 1001–1012.

Kanatsu-Shinohara, M., Inoue, K., Ogonuki, N., Miki, H., Yoshida, S., Toyokuni, S., Lee, J., Ogura, A., and Shinohara, T. (2007). Leukemia inhibitory factor enhances formation of germ cell colonies in neonatal mouse testis culture. Biol Reprod *76*, 55–62.

Kanatsu-Shinohara, M., Lee, J., Inoue, K., Ogonuki, N., Miki, H., Toyokuni, S., Ikawa, M., Nakamura, T., Ogura, A., and Shinohara, T. (2008). Pluripotency of a single spermatogonial stem cell in mice. Biol Reprod *78*, 681–687.

Kanatsu-Shinohara, M., Miki, H., Inoue, K., Ogonuki, N., Toyokuni, S., Ogura, A., and Shinohara, T. (2005a). Long-term culture of mouse male germline stem cells under serum-or feeder-free conditions. Biol Reprod *72*, 985–991.

Kanatsu-Shinohara, M., Ogonuki, N., Inoue, K., Miki, H., Ogura, A., Toyokuni, S., and Shinohara, T. (2003). Long-term proliferation in culture and germline transmission of mouse male germline stem cells. Biol Reprod *69*, 612–616.

Kanatsu-Shinohara, M., Ogonuki, N., Iwano, T., Lee, J., Kazuki, Y., Inoue, K., Miki, H., Takehashi, M., Toyokuni, S., Shinkai, Y., et al. (2005b). Genetic and epigenetic properties of mouse male germline stem cells during long-term culture. Development *132*, 4155–4163.

Kim, J., Seandel, M., Falciatori, I., Wen, D., and Rafii, S. (2008). CD34+ Testicular Stromal Cells Support Long-Term Expansion of Embryonic and Adult Stem and Progenitor Cells. Stem Cells *26*, 2516–2522.

Ko, K., Tapia, N., Wu, G., Kim, J.B., Bravo, M.J.A., Sasse, P., Glaser, T., Ruau, D., Han, D.W., Greber, B., et al. (2009). Induction of pluripotency in adult unipotent germline stem cells. Cell Stem Cell *5*, 87–96.

Kossack, N., Meneses, J., Shefi, S., Nguyen, H.N., Chavez, S., Nicholas, C., Gromoll, J., Turek, P.J., and Reijo-Pera, R.A. (2009). Isolation and characterization of pluripotent human spermatogonial stem cell-derived cells. Stem Cells *27*, 138–149.

Kota, S.K., and Feil, R. (2010). Epigenetic transitions in germ cell development and meiosis. Dev Cell *19*, 675–686.

Kouzarides, T. (2007). Chromatin Modifications and Their Function. Cell *128*, 693–705.

Kruidenier, L., Chung, C.-W., Cheng, Z., Liddle, J., Che, K., Joberty, G., Bantscheff, M., Bountra, C., Bridges, A., Diallo, H., et al. (2012). A selective jumonji H3K27 demethylase inhibitor modulates the proinflammatory macrophage response. Nature *488*, 404–408.

Kubota, H., Avarbock, M.R., and Brinster, R.L. (2003). Spermatogonial stem cells share some, but not all, phenotypic and functional characteristics with other stem cells. Proc Natl Acad Sci USA *100*, 6487–6492.

Kubota, H., Avarbock, M.R., and Brinster, R.L. (2004a). Culture Conditions and Single Growth Factors Affect Fate Determination of Mouse Spermatogonial Stem Cells. Biol Reprod *71*, 722–731.

Kubota, H., Avarbock, M.R., and Brinster, R.L. (2004b). Growth factors essential for self-renewal and expansion of mouse spermatogonial stem cells. Proc Natl Acad Sci USA *101*, 16489–16494.

Kurimoto, K., Yabuta, Y., Ohinata, Y., Shigeta, M., Yamanaka, K., and Saitou, M. (2008). Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. Genes Dev *22*, 1617–1635.

Lan, F., Bayliss, P.E., Rinn, J.L., Whetstine, J.R., Wang, J.K., Chen, S., Iwase, S., Alpatov, R., Issaeva, I., Canaani, E., et al. (2007). A histone H3 lysine 27 demethylase regulates animal posterior development. Nature *449*, 689–694.

Lesch, B.J., and Page, D.C. (2014). Poised chromatin in the mammalian germ line. Development *141*, 3619–3626.

Lesch, B.J., Dokshin, G.A., Young, R.A., McCarrey, J.R., and Page, D.C. (2013). A set of genes critical to development is epigenetically poised in mouse germ cells from fetal stages through completion of meiosis. Proceedings of the National Academy of Sciences *110*, 16061–16066.

Lewis, P.W., Elsaesser, S.J., Noh, K.-M., Stadler, S.C., and Allis, C.D. (2010). Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. Proc Natl Acad Sci USA *107*, 14075–14080.

Li, B., Carey, M., and Workman, J.L. (2007). The role of chromatin during transcription. Cell *128*, 707–719.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, J.-Y., Lees-Murdock, D.J., Xu, G.-L., and Walsh, C.P. (2004). Timing of establishment of paternal methylation imprints in the mouse. Genomics *84*, 952–960.

Loh, Y.-H., Wu, Q., Chew, J.-L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., et al. (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat Genet *38*, 431–440.

Loya, K., Eggenschwiler, R., Ko, K., Sgodda, M., André, F., Bleidißel, M., Scholer, H.R., and Cantz, T. Hepatic differentiation of pluripotent stem cells. Biological Chemistry *390*.

Loyola, A., Bonaldi, T., Roche, D., Imhof, A., and Almouzni, G. (2006). PTMs on H3 variants before chromatin assembly potentiate their final epigenetic state. Mol Cell *24*, 309–316.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251–260.

Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. Nature *469*, 343–349.

Margueron, R., Li, G., Sarma, K., Blais, A., Zavadil, J., Woodcock, C.L., Dynlacht, B.D., and Reinberg, D. (2008). Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. Mol Cell *32*, 503–518.

Mason, P.B., and Struhl, K. (2005). Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo. Mol Cell *17*, 831–840.

McKittrick, E., Gafken, P.R., Ahmad, K., and Henikoff, S. (2004). Histone H3.3 is enriched in covalent modifications associated with active chromatin. Proc Natl Acad Sci USA *101*, 1525–1530.

Meng, X., Lindahl, M., Hyvönen, M.E., Parvinen, M., de Rooij, D.G., Hess, M.W., Raatikainen-Ahokas, A., Sainio, K., Rauvala, H., Lakso, M., et al. (2000). Regulation of cell fate decision of undifferentiated spermatogonia by GDNF. Science *287*, 1489–1493.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553–560.

Mito, Y., Henikoff, J.G., and Henikoff, S. (2005). Genome-scale profiling of histone H3.3 replacement patterns. Nat Genet *37*, 1090–1097.

Mito, Y., Henikoff, J.G., and Henikoff, S. (2007). Histone replacement marks the boundaries of cis-regulatory domains. Science *315*, 1408–1411.

Moraveji, S.-F., Attari, F., Shahverdi, A., Sepehri, H., Farrokhi, A., Hassani, S.-N., Fonoudi, H., Aghdami, N., and Baharvand, H. (2012). Inhibition of glycogen synthase kinase-3 promotes efficient derivation of pluripotent stem cells from neonatal mouse testis. Hum Reprod *27*, 2312–2324.

Morrow, C.M.K., Hostetler, C.E., Griswold, M.D., Hofmann, M.-C., Murphy, K.M., Cooke, P.S., and Hess, R.A. (2007). ETV5 Is Required for Continuous Spermatogenesis in Adult Mice and May Mediate Blood Testes Barrier Function and Testicular Immune Privilege. Ann N Y Acad Sci *1120*, 144–151.

Myers, S., Bowden, R., Tumian, A., Bontrop, R.E., Freeman, C., MacFie, T.S., McVean, G., and Donnelly, P. (2010). Drive Against Hotspot Motifs in Primates Implicates the PRDM9 Gene in Meiotic Recombination. Science *327*, 876–879.

Nagano, M., Avarbock, M.R., and Brinster, R.L. (1999). Pattern and kinetics of mouse donor spermatogonial stem cell colonization in recipient testes. Biol Reprod *60*, 1429–1436.

Nakagawa, T., Nabeshima, Y.-I., and Yoshida, S. (2007). Functional identification of the actual and potential stem cell compartments in mouse spermatogenesis. Dev Cell *12*, 195–206.

Ng, J.-H., Kumar, V., Muratani, M., Kraus, P., Yeo, J.-C., Yaw, L.-P., Xue, K., Lufkin, T., Prabhakar, S., and Ng, H.-H. (2013). In Vivo Epigenomic Profiling of Germ Cells Reveals Germ Cell Molecular Signatures. Dev Cell.

Oakes, C.C., La Salle, S., Smiraglia, D.J., Robaire, B., and Trasler, J.M. (2007).

Developmental acquisition of genome-wide DNA methylation occurs prior to meiosis in male germ cells. Dev Biol *307*, 368–379.

Oatley, J.M., and Brinster, R.L. (2006). Spermatogonial stem cells. Meth Enzymol *419*, 259–282.

Oatley, J.M., Avarbock, M.R., Telaranta, A.I., Fearon, D.T., and Brinster, R.L. (2006). Identifying genes important for spermatogonial stem cell self-renewal and survival. Proc Natl Acad Sci USA *103*, 9524–9529.

Oatley, M.J., Kaucher, A.V., Racicot, K.E., and Oatley, J.M. (2011a). Inhibitor of DNA binding 4 is expressed selectively by single spermatogonia in the male germline and regulates the self-renewal of spermatogonial stem cells in mice. Biol Reprod *85*, 347–356.

Oatley, M.J., Racicot, K.E., and Oatley, J.M. (2011b). Sertoli Cells Dictate Spermatogonial Stem Cell Niches in the Mouse Testis. Biol Reprod *84*, 639–645.

Ohta, H., Tohda, A., and Nishimune, Y. (2003). Proliferation and differentiation of spermatogonial stem cells in the w/wv mutant mouse testis. Biol Reprod *69*, 1815–1821.

Parvanov, E.D., Petkov, P.M., and Paigen, K. (2010). Prdm9 controls activation of mammalian recombination hotspots. Science *327*, 835.

Payne, C., and Braun, R. (2006). Histone lysine trimethylation exhibits a distinct perinuclear distribution in Plzf-expressing spermatogonia. Dev Biol *293*, 461–472.

Pesce, M., Wang, X., Wolgemuth, D.J., and Scholer, H. (1998). Differential expression of the Oct-4 transcription factor during mouse germ cell differentiation. Mech Dev *71*, 89–98.

Phillips, B.T., Gassei, K., and Orwig, K.E. (2010). Spermatogonial stem cell regulation and spermatogenesis. Philos Trans R Soc Lond, B, Biol Sci *365*, 1663–1678.

Pittoggi, C., Renzi, L., Zaccagnini, G., Cimini, D., Degrassi, F., Giordano, R., Magnano, A.R., Lorenzini, R., Lavia, P., and Spadafora, C. (1999). A fraction of mouse sperm chromatin is organized in nucleosomal hypersensitive domains enriched in retroposon DNA. J Cell Sci *112 (Pt 20)*, 3537–3548.

Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Lee, T.I., Bell, G.W., Walker, K., Rolfe, P.A., Herbolsheimer, E., et al. (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. Cell *122*, 517–527.

Pusarla, R.-H., and Bhargava, P. (2005). Histones in functional diversification. Core histone variants. Febs J *272*, 5149–5168.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rato, L., Alves, M.G., Socorro, S., Duarte, A.I., Cavaco, J.E., and Oliveira, P.F. (2012). Metabolic regulation is important for spermatogenesis. Nat Rev Urol *9*, 330–338.

Rayasam, G., Wendling, O., Angrand, P., Mark, M., Niederreither, K., Song, L., Lerouge, T., Hager, G., Chambon, P., and Losson, R. (2003). NSD1 is essential for early post-implantation development and has a catalytically active SET domain. Embo J *22*, 3153–3163.

Roberts, C., Sutherland, H.F., Farmer, H., Kimber, W., Halford, S., Carey, A., Brickman, J.M., Wynshaw-Boris, A., and Scambler, P.J. (2002). Targeted mutagenesis of the Hira gene results in gastrulation defects and patterning abnormalities of mesoendodermal derivatives prior to early embryonic lethality. Mol Cell Biol *22*, 2318–2328.

Ruthenburg, A.J., Li, H., Patel, D.J., and Allis, C.D. (2007). Multivalent engagement of chromatin modifications by linked binding modules. Nat Rev Mol Cell Biol *8*, 983–994.

Sachs, M., Onodera, C., Blaschke, K., Ebata, K.T., Song, J.S., and Ramalho-Santos, M. (2013). Bivalent chromatin marks developmental regulatory genes in the mouse embryonic germline in vivo. Cell Rep *3*, 1777–1784.

Sadri-Ardekani, H., Mizrak, S.C., van Daalen, S.K.M., Korver, C.M., Roepers-Gajadien, H.L., Koruji, M., Hovingh, S., de Reijke, T.M., la Rosette, de, J.J.M.C.H., van der Veen, F., et al. (2009). Propagation of human spermatogonial stem cells in vitro. Jama *302*, 2127–2134.

Saitou, M., Barton, S.C., and Surani, M.A. (2002). A molecular programme for the specification of germ cell fate in mice. Nature *418*, 293–300.

Samavarchi-Tehrani, P., Golipour, A., David, L., Sung, H.-K., Beyer, T.A., Datti, A., Woltjen, K., Nagy, A., and Wrana, J.L. (2010). Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. Cell Stem Cell *7*, 64–77.

Santenard, A., Ziegler-Birling, C., Koch, M., Tora, L., Bannister, A.J., and Torres-Padilla, M.-E. (2010). Heterochromatin formation in the mouse embryo requires critical residues of the histone variant H3.3. Nat Cell Biol *12*, 853–862.

Sasaki, H., and Matsui, Y. (2008). Epigenetic events in mammalian germ-cell development: reprogramming and beyond. Nat Rev Genet *9*, 129–140.

Seandel, M., Falciatori, I., and Rafii, S. (2010). Male Germline Stem Cells: Developmental and Regenerative Potential (Totowa, NJ: Humana Press).

Seandel, M., James, D., Shmelkov, S.V., Falciatori, I., Kim, J., Chavala, S., Scherr, D.S., Zhang, F., Torres, R., Gale, N.W., et al. (2007). Generation of functional multipotent adult stem cells from GPR125+ germline progenitors. Nature *449*, 346–350.

Seki, Y., Hayashi, K., Itoh, K., Mizugaki, M., Saitou, M., and Matsui, Y. (2005). Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. Dev Biol *278*, 440–458.

Seki, Y., Yamaji, M., Yabuta, Y., Sano, M., Shigeta, M., Matsui, Y., Saga, Y., Tachibana, M., Shinkai, Y., and Saitou, M. (2007). Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice. Development *134*, 2627–2638.

Shen, X., Liu, Y., Hsu, Y.-J., Fujiwara, Y., Kim, J., Mao, X., Yuan, G.-C., and Orkin, S.H. (2008). EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. Mol Cell *32*, 491–502.

Shi, X., Hong, T., Walter, K.L., Ewalt, M., Michishita, E., Hung, T., Carney, D., Peña, P., Lan, F., Kaadige, M.R., et al. (2006). ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. Nature *442*, 96–99.

Shilatifard, A. (2008). Molecular implementation and physiological roles for histone H3 lysine 4 (H3K4) methylation. Curr Opin Cell Biol *20*, 341–348.

Shinohara, T., Avarbock, M.R., and Brinster, R.L. (1999). beta1- and alpha6-integrin are surface markers on mouse spermatogonial stem cells. Proc Natl Acad Sci USA *96*, 5504–5509.

Shinohara, T., Avarbock, M.R., and Brinster, R.L. (2000a). Functional analysis of spermatogonial stem cells in Steel and cryptorchid infertile mouse models. Dev Biol *220*, 401–411.

Shinohara, T., Orwig, K.E., Avarbock, M.R., and Brinster, R.L. (2000b). Spermatogonial stem cell enrichment by multiparameter selection of mouse testis cells. Proc Natl Acad Sci USA *97*, 8346–8351.

Shinohara, T.T., Orwig, K.E.K., Avarbock, M.R.M., and Brinster, R.L.R. (2001). Remodeling of the postnatal mouse testis is accompanied by dramatic changes in stem cell number and niche accessibility. Proc Natl Acad Sci USA *98*, 6186–6191.

Sierra, F., Lichtler, A., Marashi, F., Rickles, R., Dyke, T., Clark, S., Wells, J., Stein,

G., and Stein, J. (1982). Organization of Human Histone Genes. Proc Natl Acad Sci USA *79*, 1795–1799.

Simon, L., Ekman, G.C., Kostereva, N., Zhang, Z., Hess, R.A., Hofmann, M.-C., and Cooke, P.S. (2009). Direct transdifferentiation of stem/progenitor spermatogonia into reproductive and nonreproductive tissues of all germ layers. Stem Cells *27*, 1666–1675.

Spivakov, M., and Fisher, A.G. (2007). Epigenetic signatures of stem-cell identity. Nat Rev Genet *8*, 263–271.

Stevens, L.C., and Little, C.C. (1954). Spontaneous Testicular Teratomas in an Inbred Strain of Mice. Proc Natl Acad Sci USA *40*, 1080–1087.

Surani, M., Hayashi, K., and Hajkova, P. (2007). Genetic and Epigenetic Regulators of Pluripotency. Cell *128*, 747–762.

Szabó, P.E., Hübner, K., (null), and Mann, J.R. (2002). Allele-specific expression of imprinted genes in mouse migratory primordial germ cells. Mech Dev *115*, 157–160.

Szenker, E., Ray-Gallet, D., and Almouzni, G. (2011). The double face of the histone variant H3.3. Cell Res *21*, 421–434.

Tagami, H., Ray-Gallet, D., Almouzni, G., and Nakatani, Y. (2004). Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. Cell *116*, 51–61.

Tagelenbosch, R.A., and De Rooij, D.G. (1993). A quantitative study of spermatogonial multiplication and stem cell renewal in the C3H/101 F 1 hybrid mouse. Mutat Res *290*, 193–200.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861–872.

Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinformatics *14*, 178–192.

Torres-Padilla, M.-E., Bannister, A.J., Hurd, P.J., Kouzarides, T., and Zernicka-Goetz, M. (2006). Dynamic distribution of the replacement histone variant H3.3 in the mouse oocyte and preimplantation embryos. Int J Dev Biol *50*, 455–461.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice

junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol *28*, 516–520.

Ueda, T., Abe, K., Miura, A., Yuzuriha, M., Zubair, M., Noguchi, M., Niwa, K., Kawase, Y., Kono, T., Matsuda, Y., et al. (2000). The paternal methylation imprint of the mouse H19 locus is acquired in the gonocyte stage during foetal testis development. Genes Cells *5*, 649–659.

van der Heijden, G.W., Derijck, A.A.H.A., Pósfai, E., Giele, M., Pelczar, P., Ramos, L., Wansink, D.G., van der Vlag, J., Peters, A.H.F.M., and de Boer, P. (2007). Chromosome-wide nucleosome replacement and H3.3 incorporation during mammalian meiotic sex chromosome inactivation. Nature Publishing Group *39*, 251–258.

van der Heijden, G.W., Dieker, J.W., Derijck, A.A.H.A., Muller, S., Berden, J.H.M., Braat, D.D.M., van der Vlag, J., and de Boer, P. (2005). Asymmetry in histone H3 variants and lysine methylation between paternal and maternal chromatin of the early mouse zygote. Mech Dev *122*, 1008–1022.

Wellman, S.E., Casano, P.J., Pilch, D.R., Marzluff, W.F., and Sittman, D.B. (1987). Characterization of mouse H3.3-like histone genes. Gene *59*, 29–39.

Wen, B., Wu, H., Shinkai, Y., Irizarry, R.A., and Feinberg, A.P. (2009). Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. Nat Genet *41*, 246–250.

Wen, D., Banaszynski, L.A., Liu, Y., Geng, F., Noh, K.-M., Xiang, J., Elemento, O., Rosenwaks, Z., Allis, C.D., and Rafii, S. (2014a). Histone variant H3.3 is an essential maternal factor for oocyte reprogramming. Proceedings of the National Academy of Sciences *111*, 7325–7330.

Wen, D., Banaszynski, L.A., Rosenwaks, Z., Allis, C.D., and Rafii, S. (2014b). H3.3 replacement facilitates epigenetic reprogramming of donor nuclei in somatic cell nuclear transfer embryos. Nucleus *5*.

Wen, D., Noh, K.-M., Goldberg, A.D., Allis, C.D., Rosenwaks, Z., Rafii, S., and Banaszynski, L.A. (2014c). Genome editing a mouse locus encoding a variant histone, H3.3B, to report on its expression in live animals. Genesis.

West, J.A., Viswanathan, S.R., Yabuuchi, A., Cunniff, K., Takeuchi, A., Park, I.-H., Sero, J.E., Zhu, H., Perez-Atayde, A., Frazier, A.L., et al. (2009). A role for Lin28 in primordial germ-cell development and germ-cell malignancy. Nature *460*, 909–913.

Wong, L., Ren, H., Williams, E., McGhie, J., Ahn, S., Sim, M., Tam, A., Earle, E., Anderson, M., Mann, J., et al. (2009). Histone H3.3 incorporation provides a unique and functionally essential telomeric chromatin in embryonic stem cells. Genome Res.

Wu, X., Oatley, J.M., Oatley, M.J., Kaucher, A.V., Avarbock, M.R., and Brinster, R.L. (2010). The POU domain transcription factor POU3F1 is an important intrinsic regulator of GDNF-induced survival and self-renewal of mouse spermatogonial stem cells. Biol Reprod *82*, 1103–1111.

Wysocka, J., Swigut, T., Milne, T., Dou, Y., Zhang, X., Burlingame, A., Roeder, R., Brivanlou, A., and Allis, C. (2005). WDR5 associates with histone H3 methylated at K4 and is essential for H3K4 methylation and vertebrate development. Cell *121*, 859–872.

Yabuta, Y., Kurimoto, K., Ohinata, Y., Seki, Y., and Saitou, M. (2006). Gene expression dynamics during germline specification in mice identified by quantitative single-cell gene expression profiling. Biol Reprod *75*, 705–716.

Yamaguchi, S., Kimura, H., Tada, M., Nakatsuji, N., and Tada, T. (2005). Nanog expression in mouse germ cell development. Gene Expression Patterns *5*, 639–646.

Yamane, K., Tateishi, K., Klose, R.J., Fang, J., Fabrizio, L.A., Erdjument-Bromage, H., Taylor-Papadimitriou, J., Tempst, P., and Zhang, Y. (2007). PLU-1 is an H3K4 demethylase involved in transcriptional repression and breast cancer cell proliferation. Mol Cell *25*, 801–812.

Ye, T., Krebs, A.R., Choukrallah, M.-A., Keime, C., Plewniak, F., Davidson, I., and Tora, L. (2011). seqMINER: an integrated ChIP-seq data interpretation platform. Nucleic Acids Res *39*, e35.

Yoshida, S. (2008). Spermatogenic stem cell system in the mouse testis. Cold Spring Harb Symp Quant Biol *73*, 25–32.

Yoshinaga, K., Nishikawa, S., Ogawa, M., Hayashi, S., Kunisada, T., Fujimoto, T., and Nishikawa, S. (1991). Role of c-kit in mouse spermatogenesis: identification of spermatogonia as a specific site of c-kit expression and function. Development *113*, 689–699.

Yuen, B.T.K., Bush, K.M., Barrilleaux, B.L., Cotterman, R., and Knoepfler, P.S. (2014). Histone H3.3 regulates dynamic chromatin states during spermatogenesis. Development *141*, 3483–3494.

Zhang, Y. (2006). It takes a PHD to interpret histone methylation. Nat Struct Mol Biol *13*, 572–574.

Zheng, K., Wu, X., Kaestner, K.H., and Wang, P.J. (2009). The pluripotency factor LIN28 marks undifferentiated spermatogonia in mouse. BMC Dev Biol *9*, 38.

Zhou, V.W., Goren, A., and Bernstein, B.E. (2011). Charting histone modifications and the functional organization of mammalian genomes. Nat Rev Genet *12*, 7–18.